Contents

Original Papers

Comparing Self-Reported Running Distance and Pace With a Commercial Fitness Watch Data: Reliability Study (e39211)
Garrett Bullock, Joanne Stocks, Benjamin Feakins, Zahra Alizadeh, Amelia Arundale, Stefan Kluzek. .......................................................... 5

Prevalence of Body Dysmorphic Disorder in the Spanish Population: Cross-Sectional Web-Based Questionnaire Study (e46515)
Álvaro Loewen, Hilario Blasco-Fontecilla, Chao Li, Marcos Bella-Fernández, Belén Ruiz-Antorán. .......................................................... 14

Methodological Insights on Recruitment and Retention From a Remote Randomized Controlled Trial Examining the Effectiveness of an Alcohol Reduction App: Descriptive Analysis Study (e51839)
Melissa Oldham, Larisa Dinu, Gemma Loebenberg, Matt Field, Matthew Hickman, Susan Michie, Jamie Brown, Claire Garnett. ......................... 29

Remote Delivery of the Cuidándome Telehealth Intervention for Self-Management of Depression and Anxiety Among Latina Immigrant Women: Randomized Controlled Trial (e52969)
Carmen Alvarez, Subhash Aryal, Elizabeth Vrany, Maria Sanchez R, Rosalphtie Quiles, Lia Escobar-Acosta, Felicia Hill-Briggs. .......................... 42

Feasibility and Acceptability of a Mobile Technology Intervention to Support Postabortion Care After Surgical Abortion (the FACTS Study Phase 3): Mixed Methods Prospective Pilot Study (e46284)
Roopan Gill, Gina Ogilvie, Wendy Norman, Brian Fitzsimmons, Ciara Maher, Regina Renner. .......................................................... 58

The Sukaribit Smartphone App for Better Self-Management of Type 2 Diabetes: Randomized Controlled Feasibility Study (e46222)
Cecilia Josefsson, Thea Liljeroos, Margareta Hellgren, Ulrika Pöder, Mariann Hedström, Erik Olsson. .......................................................... 72

Office Workers’ Views About the Uses, Concerns, and Acceptance of Hand Hygiene Data Collected From Smart Sanitizers: Exploratory Qualitative Interview Study (e47308)
Sophie Rutter, Sally Sanger, Andrew Madden, Sukaina Ehdeed, Catherine Stones. .......................................................... 87

Population-Level Portal-Based Anxiety and Depression Screening Perspectives in HIV Care Clinicians: Qualitative Study Using the Consolidated Framework for Implementation Research (e48935)
Daniela Zimmer, Erin Staab, Jessica Ridgway, Jessica Schmitt, Melissa Franco, Scott Hunter, Darnell Motley, Neda Laiteerapong. ......................... 100

Dentists’ Information Needs and Opinions on Accessing Patient Information via Health Information Exchange: Survey Study (e51200)
Shuning Li, Grace Felix Gomez, Huiping Xu, Anushri Rajapuri, Brian Dixon, Thankam Thyvalikakath. .......................................................... 112
Promoting Self-Efficacy of Individuals With Autism in Practicing Social Skills in the Workplace Using Virtual Reality and Physiological Sensors: Mixed Methods Study (e52157)
Sung-In Kim, So-youn Jang, Taewan Kim, Bogoan Kim, Davyoung Jeon, Taehyung Noh, Minkyu Jeong, Kaely Hal, Meelim Kim, Hee Yoo, Kyung Sik Han, Hwajung Hong, Jennifer Kim. ................................................................. 124

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

Linguistic Variables and Gender Differences Within a Messenger-Based Psychosocial Chat Counseling Service for Children and Adolescents: Cross-Sectional Study (e51795)
Zeki Efe, Sabrina Baldofski, Elisabeth Kohls, Melanie Eckert, Shadi Saee, Julia Thomas, Richard Wundrack, Christine Rummel-Kluge. .................................................. 149

Rosie, a Health Education Question-and-Answer Chatbot for New Mothers: Randomized Pilot Study (e51361)
Quynh Nguyen, Elizabeth Aparicio, Michelle Jasczynski, Amara Channell Doig, Xiaohe Yue, Heran Mane, Neha Srikanth, Francis Gutierrez, Nataly Delcid, Xin He, Jordan Boyd-Graber. ................................................................. 164

Clinical Needs Assessment of a Machine Learning–Based Asthma Management Tool: User-Centered Design Approach (e45391)
Lu Zheng, Joshua Ohde, Shauna Overgaard, Tracey Brereton, Kristelle Jose, Chung-Il Wi, Kevin Peterson, Young Juhn. ................................................................. 177

Exploring Adolescents’ Attitudes Toward Mental Health Apps: Concurrent Mixed Methods Study (e50222)
Helene Høgsdal, Henriette Kyrrestad, Marte Rye, Sabine Kaiser. ................................................................. 188

The Impact of Social Media Use on Mental Health and Family Functioning Within Web-Based Communities in Saudi Arabia: Ethnographic Correlational Study (e44923)
Bdour Alwuqaysi, Allie Abdul-Rahman, Rita Borgo. ................................................................. 198

Enabling Personalization for Digital Cognitive Stimulation to Support Communication With People With Dementia: Pilot Intervention Study as a Prelude to AI Development (e51732)
Nick Hird, Tohmi Osaki, Samik Ghosh, Sucheendr Palaniappan, Kiyoshi Maeda. ................................................................. 213

Designing and Developing a Mobile App for Management and Treatment of Gestational Diabetes in Nepal: User-Centered Design Study (e50823)
Aarthi Shanmugavel, Prabin Shakya, Archana Shrestha, Jyoti Nepal, Abha Shrestha, Jean-Francois Daneault, Shristi Rawal. ................................................................. 231

Preoperative Virtual Reality to Expose Patients With Breast Cancer to the Operating Room Environment: Feasibility and Pilot Case Series Study (e46367)
Jordana Sommer, Kristin Reynolds, Pamela Hebbard, Michael Smith, Natalie Mota, W Mutch, Jessica Maples-Keller, Leslie Roos, Renée El-Gabalawy. ................................................................. 248

Educational Video Intervention to Improve Health Misinformation Identification on WhatsApp Among Saudi Arabian Population: Pre-Post Intervention Study (e50211)
Ebtithal Alsaaad, Sharifah AlDossary. ................................................................. 265

A Blended Intervention Targeting Emotion Dysregulation in Adults With Attention-Deficit/Hyperactivity Disorder: Development and Feasibility Study (e53931)
Emilie Nordby, Frode Guribye, Viktor Schenning, Sander Andersen, Jonna Kunts, Astrid Lundervold. ................................................................. 280

Designing a Web-Based Navigation Tool to Support Access to Youth Mental Health Services: Qualitative Study (e48945)
Alison Calear, Philip Batterham, Sonia McCallum, Michelle Banfield, Elizabeth Moore, Natalie Johnson, Alyssa Morse. ................................................................. 296
Development and Usability Testing of an mHealth Tool for Trauma-Informed Prevention of Substance Use, HIV Acquisition, and Risky Sexual Behaviors Among Adolescents: Mixed Methods Study (e52835)
Carla Danielson, Angela Moreland, Austin Hahn, Devin Banks, Kenneth Ruggiero. .......................................................... 307

Virtual Exercise in Medicine: A Proof of Concept in a Healthy Population (e45637)
Barbara Le Roy, Charles Martin-Krumm, Charlotte Poupon, Raphaëlle Richieri, Eric Malbos, Fanny Barthélémy, Eric Guedj, Marion Trousselard. .......................................................... 317

Experiences Among Patients With Cystic Fibrosis in the MucoExocet Study of Using Connected Devices for the Management of Pulmonary Exacerbations: Grounded Theory Qualitative Research (e38064)
Maxime Morsa, Amélie Perrin, Valérie David, Gilles Rault, Enora Le Roux, Corinne Alberti, Rémi Gagnayre, Dominique Pougeon Bertrand. .......................................................... 337

Culturally Adapting the World Health Organization Digital Intervention for Family Caregivers of People With Dementia (iSupport): Community-Based Participatory Approach (e46941)
Anna Messina, Rebecca Amati, Anna Annoni, Beatrice Bano, Emiliano Albanese, Maddalena Fiordelli. .......................................................... 357

A Bluetooth-Enabled Device for Real-Time Detection of Sitting, Standing, and Walking: Cross-Sectional Validation Study (e47157)
Reza Daryabegyi-Khodbehsara, Jonathan Rawstorn, David Dunstan, Sheikh Shariful Islam, Mohamed Abdelrazek, Abbas Kouzani, Poolith Thummala, Jenna McVicar, Ralph Maddison. .......................................................... 372

Demographics and Social Factors Associated With Persistent Nonuse of Video Appointments at a Multisite Health Care Institution: Cross-Sectional Study (e50572)
Pravesh Sharma, Celia Kamath, Tabetha Brockman, Anne Roche, Pamela Sinicrope, Ruoxiang Jiang, Paul Decker, Vanessa Paudernik, Christi Patten. .......................................................... 386

Transfer of Knowledge on Pneumoconiosis Care Among Rural-Based Members of a Digital Community of Practice: Cross-Sectional Study (e52414)
Brian Soller, Orrin Myers, Akshay Sood. .......................................................... 401

Medical Students’ Perceptions on Identifying and Addressing Emotional Responses in Emergency Medicine: Pilot Investigation (e50827)
Anish Agarwal, Rachel Gonzales, Cory Munden, DaCarla Albright, Suzana Tsao. .......................................................... 412

Machine Learning for Early Prediction of Major Adverse Cardiovascular Events After First Percutaneous Coronary Intervention in Patients With Acute Myocardial Infarction: Retrospective Cohort Study (e48487)
Pin Zhang, Lei Wu, Ting-Ting Zou, ZiXuan Zou, JiaXin Tu, Ren Gong, Jie Kuang. .......................................................... 419

Differences in Psychological Inflexibility Among Men With Erectile Dysfunction Younger and Older Than 40 Years: Web-Based Cross-Sectional Study (e45998)
Junichi Saito, Hiroaki Kumano, Mohammad Ghazizadeh, Chigusa Shimokawa, Hideki Tanemura. .......................................................... 430

Determining Distinct Suicide Attempts From Recurrent Electronic Health Record Codes: Classification Study (e46364)
Kate Bentley, Emily Madsen, Eugene Song, Yu Zhou, Victor Castro, Hyunjoon Lee, Younga Lee, Jordan Smoller. .......................................................... 438

Synergizing the Behavior Change Wheel and a Cocreative Approach to Design a Physical Activity Intervention for Adolescents and Young Adults With Intellectual Disabilities: Development Study (e51693)
Laura Maenhout, Julie Latomme, Greet Cardon, Geert Crombez, Geert Van Hove, Sofie Compernolle. .......................................................... 448

Assessing and Improving Data Integrity in Web-Based Surveys: Comparison of Fraud Detection Systems in a COVID-19 Study (e47091)
Stephen Bonett, Willey Lin, Patrina Sexton Topper, James Wolfe, Jesse Golinkoff, Aayushi Deshpande, Antonia Villarruel, José Bauermeister. .......................................................... 467
Virtual and Interprofessional Objective Structured Clinical Examination in Dentistry and Dental Technology: Development and User Evaluations (e44653)
MengWei Pang, YanLing Dong, XiaoHan Zhao, JiaWu Wan, Li Jiang, JinLin Song, Ping Ji, Lin Jiang .......................... 481

Use of Machine Learning Tools in Evidence Synthesis of Tobacco Use Among Sexual and Gender Diverse Populations: Algorithm Development and Validation (e49031)
Shaoying Ma, Shuning Jiang, Olivia Yang, Xuanzhi Zhang, Yu Fu, YuSen Zhang, Aadeeba Kaareen, Meng Ling, Jian Chen, Ce Shang .......................... 505

Acceptance of Medical Artificial Intelligence in Skin Cancer Screening: Choice-Based Conjoint Survey (e46402)
Inga Jagemann, Ole Wensing, Manuel Stegemann, Gerrit Hirschedler ................................................................. 517

Digital Health Needs and Preferences During Pregnancy and the Postpartum Period: Mixed Methods Study (e48960)
Natalie Henrich, Alison Brinson, Alyssa Arnold, Hannah Jahnke ................................................................. 527

Acceptability and Feasibility of a Smartphone-Based Real-Time Assessment of Suicide Among Black Men: Mixed Methods Pilot Study (e48992)
Leslie Adams, Thomasina Watts, Aubrey DeVinney, Emily Haroz, Johannes Thrul, Jasmin Stephens, Mia Campbell, Denis Antoine, Benjamin Lé Cook, Sean Joe, Roland Thorpe Jr .......................... 545

Privacy Concerns About Sharing General and Specific Health Information on Twitter: Quantitative Study (e45573)
Pouyan Esmaeilzadeh ................................................................. 557

Assessment of Qatar’s Health Care Community Call Center Efficacy in Addressing COVID-19 Pandemic Health Care Challenges: Cross-Sectional Study (e42753)
Muhammad Waheed, Lolwa Al Mannai, Hanan Khudadad, Jamil Alenbawi, Mariama Mansaray, Samya Al Abdulla .......................... 580

Web-Based Search Volume for HIV Tests and HIV-Testing Preferences During the COVID-19 Pandemic in Japan: Infodemiology Study (e52306)
Rie Kanamori, Futaba Uemura, Kosuke Uemura, Taiju Miyagami, Simon Valenti, Nobuyuki Fukui, Mayumi Yuda, Mizue Saita, Hirotake Mori, Toshio Naito ................................................................. 594

Corrigenda and Addenda

Correction: Evaluating the Clinical Feasibility of an Artificial Intelligence–Powered, Web-Based Clinical Decision Support System for the Treatment of Depression in Adults: Longitudinal Feasibility Study (e56570)
Christina Popescu, Grace Golden, David Benrimoh, Myriam Tanguay-Sela, Dominique Slowey, Eryn Lundrigan, Jérôme Williams, Bennet Desormeau, Divyesh Kardani, Tamara Perez, Collen Rollins, Sonia Israel, Kelly Perlman, Caitrin Armstrong, Jacob Baxter, Kate Whitmore, Marie-Jeanne Fradette, Kaelan Felcarek-Hope, Ghassen Souli, Robert Fratila, Joseph Mehlitter, Karl Looper, Warren Steiner, Soham Rej, Jordan Karp, Katherine Heller, Sagar Parikh, Rebecca McGuire-Snedkus, Manuela Ferrari, Howard Margoless, Gustavo Turecki .......................... 592
Comparing Self-Reported Running Distance and Pace With a Commercial Fitness Watch Data: Reliability Study

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

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

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

Abstract

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

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

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

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

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

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

KEYWORDS

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

Introduction

Physical activity is an essential component of a healthy lifestyle [1]. There is a substantial body of evidence highlighting the physical, social, and psychological health benefits of regular physical activity [1-3]. Sustainable physical activity interventions are needed, given that 31% of the global population is sedentary [4]. The World Health Organization’s physical activity action plan [5] identifies sport as an underused yet significant contributor to physical activity.
One widely popular sport globally is running [6]. Over the past 40 years, running has become one of the most popular physical leisure activities [7,8]. An estimated 50 million people in Europe participate in running as a way to stay healthy [9]. Due to high running participation prevalence [9], researchers have attempted to quantify running habits and training load, most notably through self-report [10]. Running load or workload is the distance run in 1 session. A training session is 1 running bout. Running speed is the intensity at which one runs for 1 running session [9,10]. However, there are potential inaccuracies from over–self-reporting due to recall bias [11] and social desirability of higher levels of physical activity [12], with potential differences by sex and age groups [13]. Further, the reliability of self-reported running pace has not been investigated, which is an important factor in quantifying running training intensity [14]. Due to these issues, research has investigated the reliability of wearables in quantifying running load [15]. Wearables, such as accelerometers, have demonstrated excellent reliability in assessing gait patterns, acceleration, and velocity [15].

Although wearable accelerometers are ubiquitously used in the general population [16] and are reliably used in research to measure physical activity levels [17], they are rarely used by running populations to track running load and training [18]. Runners opt for wearable GPS watches to track running training [19], with up to 90% of regular runners using some form of GPS monitoring when running [18]. GPS wearable technology quantifies running workload and speed [20]. A systematic review determined that there is excellent reliability for step counting and moderate validity for energy expenditure and distance run [21]. The most popular GPS wearable technology used by runners is the Garmin watch, as indicated by a previous survey where 44% of GPS and sports watches were Garmin, compared to 27% for Polar and 7% for Nike watches [22].

There has been previous related work in evaluating the reliability of running self-reports in large samples. In a sample of 92 endurance runners, followed for a 52-week (ie, 1 year) period, 93% of the runners participated in the entire follow-up period [10]. In a study of 53 running participants over 18 weeks, the response rate was 73% over the reporting period [23]. Another study surveyed 228 runners for at least 6 months, with a 2.2% attrition rate [24].

Although there is substantial evidence investigating the reliability of running distance self-reporting [10] and GPS wearable technology [20], there are currently no studies investigating the reliability of participant self-reporting in comparison to GPS wearable technology in running populations. There is a critical sports science and medical research gap due to a paucity of reliability studies assessing self-report running pace. Further, as GPS wearable technology is expensive [25], there may be a barrier for some recreational runners, decreasing the efficacy of using GPS monitoring alone to assess running workload [22]. Therefore, this study aimed to assess the reliability of weekly self-reported running distance and pace compared to a commercial GPS fitness watch, stratified by sex and age. These data will give clinicians and sports researchers insights into the reliability of runners’ self-reported pace, which may improve training designs and rehabilitation prescriptions.

**Methods**

**Study Design**

A prospective cohort study was conducted using a mobile-based app. Participants accessed a dynamic digital consent form through the app or the recruitment website during the spring of 2021 over a 4-month period. During consent, participants could select different levels of study engagement. All levels of engagement involved the following: (1) an acknowledgement, understanding, and consent to participate in the study; (2) a baseline questionnaire collecting information on demographics, previous and current injury and illness history, footwear and foot posture, knee symptoms, lifestyle, and previous year’s training load; and (3) weekly reports on training load and each participant’s perceptions of cardiorespiratory symptoms, mood, and incidence of illness and injury in the last week. More advanced participation involved connecting participants’ Garmin Connect (Garmin Ltd) data, which included sharing data on running distance, running speed, and heart rate during each training or racing session. Participants added their Garmin Connect information at study recruitment. Garmin data were then automatically uploaded every week when the participant was within the study. Once the participant reaches the end of the study data collection or voluntarily leaves the study, the Garmin data collection link is terminated, ending data upload (Figure 1). Participants could opt out of the study at any time.
Ethical Considerations
This study received a favorable ethical review from the University of Nottingham (FMHS 113-1120). All methods were performed following the relevant guidelines and regulations of the Declaration of Helsinki. Before study inclusion, all participants were detailed about the risks and benefits of participation. All participants provided informed consent to participate.

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

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

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

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

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

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

### Results

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

#### Table 1. Participant descriptive statistics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Overall (n=475)</th>
<th>Female participants (n=248)</th>
<th>Male participants (n=227)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>49.5 (12.2)</td>
<td>51.0 (13.1)</td>
<td>47.8 (10.9)</td>
</tr>
<tr>
<td>BMI (kg/m$^2$), mean (SD)</td>
<td>24.0 (3.7)</td>
<td>24.2 (3.7)</td>
<td>23.8 (3.8)</td>
</tr>
<tr>
<td>Number of weeks followed, median (IQR)</td>
<td>17 (11-23)</td>
<td>15 (10-20)</td>
<td>17 (11-24)</td>
</tr>
<tr>
<td>Smoking, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current smoker</td>
<td>10 (2)</td>
<td>6 (2)</td>
<td>4 (1)</td>
</tr>
<tr>
<td>Ex-smoker</td>
<td>62 (13)</td>
<td>30 (13)</td>
<td>32 (13)</td>
</tr>
<tr>
<td>Cigarettes per day, median (IQR)</td>
<td>9 (3-14)</td>
<td>7 (5-10)</td>
<td>30 (23-37)</td>
</tr>
<tr>
<td>Years smoked, median (IQR)</td>
<td>22 (9-35)</td>
<td>14 (2-26)</td>
<td>25 (18-31)</td>
</tr>
<tr>
<td>Diabetes, n (%)</td>
<td>6 (1)</td>
<td>1 (1)</td>
<td>5 (3)</td>
</tr>
<tr>
<td>Heart disease, n (%)</td>
<td>5 (1)</td>
<td>2 (1)</td>
<td>3 (1)</td>
</tr>
<tr>
<td>Cancer, n (%)</td>
<td>14 (3)</td>
<td>10 (4)</td>
<td>4 (2)</td>
</tr>
<tr>
<td>Asthma, n (%)</td>
<td>62 (13)</td>
<td>40 (16)</td>
<td>32 (11)</td>
</tr>
<tr>
<td>Hay fever (pollen allergies), n (%)</td>
<td>180 (38)</td>
<td>99 (40)</td>
<td>81 (35)</td>
</tr>
<tr>
<td>Days of running per week, mean (SD)</td>
<td>2 (1)</td>
<td>2 (1)</td>
<td>2 (1)</td>
</tr>
</tbody>
</table>

#### Table 2. Weekly running descriptive characteristics using Garmin watch data.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Overall (n=475)</th>
<th>Female participants (n=248)</th>
<th>Male participants (n=227)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time running (min), median (IQR)</td>
<td>144 (48-238)</td>
<td>139 (43-226)</td>
<td>153 (45-230)</td>
</tr>
<tr>
<td>Calories burned (kcal), median (IQR)</td>
<td>1450 (547-2353)</td>
<td>1212 (461-1963)</td>
<td>1755 (671-2840)</td>
</tr>
<tr>
<td>Kilometers, median (IQR)</td>
<td>25.9 (4.7-41.8)</td>
<td>22.7 (4.5-40.9)</td>
<td>29.6 (4.9-40.6)</td>
</tr>
<tr>
<td>Running pace (min/km), median (IQR)</td>
<td>6.1 (5.2-7.0)</td>
<td>6.7 (5.7-7.7)</td>
<td>5.7 (4.9-6.4)</td>
</tr>
<tr>
<td>Average heart rate (bpm$^a$), mean (SD)</td>
<td>130 (26)</td>
<td>127 (26)</td>
<td>130 (24)</td>
</tr>
<tr>
<td>Maximum heart rate (bpm$^a$), mean (SD)</td>
<td>163 (24)</td>
<td>162 (25)</td>
<td>163 (24)</td>
</tr>
</tbody>
</table>

$^a$Bpm: beats per minute.

### Reliability

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

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

aICC: intraclass correlation coefficient.

### Discussion

#### Principal Findings

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

#### Comparison to Previous Work

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

However, this is only speculative, and further research is required.

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

These findings necessitate future research. Participants were recreational runners, and most of them were older than 40 years. Future work is needed to assess the reliability of self-report in comparison to GPS monitoring data in elite runners of all ages and younger populations across different skill or competition levels. All runners in this study already owned a Garmin watch before the study enrollment. Understanding how self-reporting changes among new GPS activity monitor users is needed.

Running pace demonstrated moderate reliability in this recreational runner sample. Future research is required to investigate the effectiveness of running pace training on self-report reliability.
Limitations
As with all studies, there are limitations to this study. First, there is the risk of recall bias due to the weekly intervals for self-reporting, which decreases the precision of these findings. Participants may not have worn or activated their Garmin watch for specific runs, decreasing the reliability of these data. GPS monitoring devices are expensive, causing a high barrier to entry. Such a barrier may add selection bias to this reliability study, decreasing the generalizability of these results to all running populations. Further, the sample in this study comprised recreational runners; therefore, the results are not generalizable to elite runners or populations that solely engage in walking for exercise. Finally, participants used different versions of the Garmin watch. As different watch versions may exhibit different reliability, there is a potential for decreased data precision.

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

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

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

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

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

Conflicts of Interest
None declared.

Multimedia Appendix 1
Study flowchart.
[DOCX File, 25 KB - formative_v8i1e39211_app1.docx ]

References


18. Mallula C. Comparing Garmin Forerunner 405CX GGP and Nike + iPod to Accurately Measure Energy Expenditure, Distance, and Speed of Overground Running. MSL Academic Endeavors. 2016. URL: https://agedscholarship.csuohio.edu/cgi/viewcontent.cgi?article=1726&context=etdarchive [accessed 2023-12-14]


24. Running through study. University of Nottingham. URL: https://www.nottingham.ac.uk/research/groups/crsm/running-through-study.aspx?text=The%20Running%20Through%20research%20team.load%2C%20intensity%20or%20infection%20or%20recovery [accessed 2021-01-01]


35. Open Science Framework. URL: https://osf.io/jgfau/ [accessed 2023-12-15]

Abbreviations

ACC: intraclass correlational coefficient
RedCAP: Research Electronic Data Capture

Edited by A Mavragani; submitted 02.05.22; peer-reviewed by K Schwartkopf, G Alce, M Pobiruchin; comments to author 30.07.22; revised version received 31.08.22; accepted 19.09.22; published 04.01.24.

Please cite as:
Comparing Self-Reported Running Distance and Pace With a Commercial Fitness Watch Data: Reliability Study
JMIR Form Res 2024;8:e39211
URL: https://formative.jmir.org/2024/1/e39211
doi:10.2196/39211
PMID:38175696

©Garrett Bullock, Joanne Stocks, Benjamin Feakins, Zahra Alizadeh, Amelia Arundale, Stefan Kluzek. Originally published in JMIR Formative Research (https://formative.jmir.org), 04.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Prevalence of Body Dysmorphic Disorder in the Spanish Population: Cross-Sectional Web-Based Questionnaire Study

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

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

Corresponding Author:
Hilario Blasco-Fontecilla, MD, PhD
Servicio de Psiquiatría Infanto-juvenil
Hospital Universitario Puerta de Hierro Majadahonda
Instituto de Investigación Sanitaria Puerta de Hierro - Segovia de Arana
C/ Manuel de Falla, 1
Majadahonda, 28222
Spain
Phone: 34 91911690
Email: hmblasco@yahoo.es

Abstract

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

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

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

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

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

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

Introduction

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The prevalence of BDD in the population assessed in this study was 15.2% (284/1421, 19.9% in females vs 34/653, 5.2% in males). Regarding age, the prevalence of BDD was higher in the youngest age group (18-24 years; 267/1091, 24.5%), followed by the 25-44 years age group (30/279, 10.8%) (Figure 1). The number of body parts of concern in participants with BDD was significantly higher than that in participants without BDD (4.6 vs 2.2, respectively; $P<.001$). Regarding the body parts of concern, body fat was the most common concern in both groups, followed by thighs, face, hips, and skin in the BDD group and thighs, teeth, and hair in the non-BDD group (Figure 2). Regarding the sociodemographic characteristics of the BDD population (n=322), the majority were females (284/322, 88.2%), with a mean age of 23.5 (SD 9.6) years, and Caucasians (243/722, 75.5%). Approximately 81.7% (1533/2091) were students, and 76.6% (247/322) of them had a university education level. Among the participants diagnosed with BDD, 63.4% (204/322) had a history of dermatologic disease, the most frequent being acne (115/204, 56.5%) and dermatitis (99/204, 48.5%). Approximately 46.6% (150/322) of the participants with BDD reported a history of psychiatric comorbidities, and the most frequent were anxiety disorders (108/150, 72%), depressive disorders (84/150, 56%), eating disorders (72/150, 48%), and attention-deficit/hyperactivity disorder (18/150, 12%). Approximately 17.7% (57/322) of the population with BDD had previously undergone a cosmetic procedure, the most frequent being laser treatment for acne, blemishes, and other skin lesions (17/57, 29.8%); mesotherapy (9/57, 15.8%); and rhinoplasty (7/57, 12.3%). The factors related to BDD are reported in univariate and multivariate analyses in Table 2. BDD diagnosis was significantly associated with female gender, other genders, age 18-24 years, students, monthly income level of less than €500 (€1=US $1.11), and participants with dermatologic and psychiatric comorbidities ($P<.001$). All these variables were included in the multivariate model, with gender (female and other), age, student occupation, depressive disorder, eating disorders, and anxiety disorders remaining at <.05 significance as diagnostic predictors of BDD.
<p>| Table 1. Sociodemographic and clinical characteristics of the participants.  |
|---------------------------------------------------------------|---------------------------------------------------------------|---------------------------------------------------------------|
|                                                               | All (N=2091)                                                  | With BDD diagnosis (n=322)                                     | Without BDD diagnosis (n=1769)                                  |
| Gender (female), n (%)                                        | 1421 (68)                                                    | 284 (88.2)                                                    | 1137 (64.3)                                                    |
| Age (years), mean (SD)                                       | 37.7 (16.6)                                                  | 23.5 (9.6)                                                    | 35.5 (16.9)                                                    |
| Age group (years), n (%)                                     |                                                               |                                                               |                                                               |
| 18-24                                                         | 1091 (52.2)                                                  | 267 (82.9)                                                    | 824 (46.6)                                                    |
| 25-44                                                         | 279 (13.3)                                                   | 30 (9.3)                                                      | 249 (14.1)                                                    |
| 45-64                                                         | 654 (31.3)                                                   | 23 (7.1)                                                      | 631 (35.7)                                                    |
| &gt;64 years                                                    | 66 (3.2)                                                     | 2 (0.6)                                                       | 64 (3.6)                                                      |
| Ethnicity, n (%)                                             |                                                               |                                                               |                                                               |
| African                                                      | 11 (0.5)                                                     | 3 (0.9)                                                       | 8 (0.5)                                                       |
| American                                                     | 6 (0.3)                                                      | 0 (0)                                                         | 6 (0.3)                                                       |
| Asian                                                        | 26 (1.2)                                                     | 7 (2.2)                                                       | 19 (1.1)                                                      |
| Caucasian                                                    | 1533 (73.3)                                                  | 243 (75.5)                                                    | 1290 (72.9)                                                   |
| Latin                                                        | 306 (14.6)                                                   | 23 (7.1)                                                      | 283 (16)                                                     |
| Other                                                        | 209 (10)                                                     | 46 (14.3)                                                    | 163 (9.2)                                                     |
| Educational level, n (%)                                     |                                                               |                                                               |                                                               |
| Elementary school                                           | 11 (0.5)                                                     | 1 (0.3)                                                       | 10 (0.6)                                                      |
| Middle school                                                | 8 (0.4)                                                      | 0 (0)                                                         | 8 (0.5)                                                       |
| Professional education                                      | 100 (4.8)                                                    | 15 (4.7)                                                      | 85 (4.8)                                                      |
| High school                                                  | 265 (12.7)                                                   | 59 (18.3)                                                    | 206 (11.6)                                                    |
| University                                                   | 1707 (81.6)                                                  | 247 (76.7)                                                    | 1460 (82.5)                                                   |
| Occupation, n (%)                                            |                                                               |                                                               |                                                               |
| Student                                                      | 1043 (49.9)                                                  | 263 (81.7)                                                    | 780 (44.1)                                                    |
| Worker                                                       | 865 (41.4)                                                   | 50 (15.5)                                                    | 815 (46.1)                                                    |
| Other                                                        | 183 (8.8)                                                    | 9 (2.8)                                                       | 174 (9.8)                                                    |
| Monthly income (€), n (%)                                    |                                                               |                                                               |                                                               |
| &lt;500                                                         | 861 (41.2)                                                   | 205 (63.7)                                                    | 656 (37.1)                                                    |
| 500-999                                                      | 138 (6.6)                                                    | 34 (10.6)                                                    | 104 (5.9)                                                      |
| 1000-1999                                                    | 343 (16.4)                                                   | 39 (12.1)                                                    | 304 (17.2)                                                    |
| 2000-2999                                                    | 291 (13.9)                                                   | 9 (2.8)                                                       | 282 (15.9)                                                    |
| &gt;3000                                                        | 358 (17.1)                                                   | 11 (3.4)                                                      | 347 (19.6)                                                    |
| Not known                                                    | 100 (4.8)                                                    | 24 (7.5)                                                      | 76 (4.3)                                                      |
| Dermatologic disease, n (%)                                  |                                                               |                                                               |                                                               |
| Acne                                                         | 1072 (51.3)                                                  | 204 (63.4)                                                    | 868 (49.1)                                                    |
| Atopic dermatitis                                           | 532 (49.6)                                                   | 115 (56.5)                                                    | 417 (48)                                                      |
| Other dermatitis                                            | 434 (40.5)                                                   | 99 (48.5)                                                    | 335 (38.6)                                                    |
| Psoriasis                                                    | 35 (3.3)                                                     | 10 (4.9)                                                      | 25 (2.9)                                                      |
| Rosacea                                                     | 50 (4.7)                                                     | 2 (1)                                                         | 48 (5.5)                                                      |
| Urticaria                                                    | 64 (6)                                                       | 5 (2.5)                                                      | 59 (6.8)                                                      |
| Pityriasis                                                   | 15 (1.4)                                                     | 1 (0.5)                                                      | 14 (1.6)                                                      |
| Eczema                                                       | 37 (3.5)                                                     | 9 (4.4)                                                      | 28 (3.2)                                                      |
| Skin infections                                              | 32 (3)                                                       | 6 (2.9)                                                      | 26 (3)                                                       |</p>
<table>
<thead>
<tr>
<th>Category</th>
<th>All (N=2091)</th>
<th>With BDD(^a) diagnosis (n=322)</th>
<th>Without BDD diagnosis (n=1769)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neoplasms</td>
<td>11 (1)</td>
<td>0 (0)</td>
<td>11 (1.3)</td>
</tr>
<tr>
<td>Other</td>
<td>65 (6.1)</td>
<td>7 (3.4)</td>
<td>58 (6.7)</td>
</tr>
<tr>
<td><strong>Psychiatric disorder, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depressive Disorder</td>
<td>493 (23.6)</td>
<td>150 (46.6)</td>
<td>343 (19.4)</td>
</tr>
<tr>
<td>Borderline personality disorder</td>
<td>244 (49.5)</td>
<td>84 (56)</td>
<td>160 (46.6)</td>
</tr>
<tr>
<td>Eating disorder</td>
<td>9 (1.8)</td>
<td>3 (2)</td>
<td>6 (1.7)</td>
</tr>
<tr>
<td>Obsessive-compulsive disorder</td>
<td>33 (6.7)</td>
<td>12 (8)</td>
<td>21 (6.1)</td>
</tr>
<tr>
<td>Attention-deficit/hyperactivity disorder</td>
<td>59 (12)</td>
<td>18 (12)</td>
<td>41 (12)</td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>7 (1.4)</td>
<td>1 (0.7)</td>
<td>6 (1.7)</td>
</tr>
<tr>
<td>Anxiety disorder</td>
<td>292 (59.2)</td>
<td>108 (72)</td>
<td>184 (53.6)</td>
</tr>
<tr>
<td>Schizophrenia/psychosis</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Substance abuse</td>
<td>12 (2.4)</td>
<td>3 (2)</td>
<td>9 (2.6)</td>
</tr>
<tr>
<td>Other</td>
<td>18 (3.7)</td>
<td>3 (2)</td>
<td>15 (4.4)</td>
</tr>
<tr>
<td>History of plastic surgery procedures, n (%)</td>
<td>355 (17)</td>
<td>57 (17.7)</td>
<td>298 (16.8)</td>
</tr>
</tbody>
</table>

\(^a\)BDD: body dysmorphic disorder.

\(^b\)€1=US $1.11.

**Figure 1.** Prevalence of body dysmorphic disorder among subsamples. BDD: body dysmorphic disorder.
Figure 2. Body parts of concern in our study population.
### Table 2. Factors related to body dysmorphic disorder.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Univariate risk ratio (95% CI)</th>
<th>P value</th>
<th>Multivariate risk ratio (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>4.55 (3.14-6.58)</td>
<td>.001</td>
<td>3.02 (2.01-4.53)</td>
<td>.001</td>
</tr>
<tr>
<td>Male</td>
<td>Ref</td>
<td></td>
<td>1.00 (1.00-1.00)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>5.60 (1.73-18.09)</td>
<td>.004</td>
<td>4.57 (1.28-16.38)</td>
<td>.02</td>
</tr>
<tr>
<td>Mean age</td>
<td>0.93 (0.92-0.95)</td>
<td>.001</td>
<td>0.97 (0.95-0.99)</td>
<td>.02</td>
</tr>
<tr>
<td><strong>Age categories (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>2.69 (1.79-4.02)</td>
<td>.001</td>
<td>1.32 (0.82-2.12)</td>
<td></td>
</tr>
<tr>
<td>25-44</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-64</td>
<td>0.30 (0.17-0.53)</td>
<td>.01</td>
<td>0.28 (0.15-0.54)</td>
<td></td>
</tr>
<tr>
<td>&gt;64</td>
<td>0.26 (0.060-1.11)</td>
<td>.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Occupation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>1.79 (1.40-2.29)</td>
<td>.001</td>
<td>1.30 (1.01-1.65)</td>
<td>.02</td>
</tr>
<tr>
<td>Worker</td>
<td>0.18 (0.13-0.25)</td>
<td>.001</td>
<td>0.51 (0.28-0.94)</td>
<td>.03</td>
</tr>
<tr>
<td>Other</td>
<td>0.15 (0.07-0.30)</td>
<td>.001</td>
<td>0.35 (0.14-0.88)</td>
<td>.03</td>
</tr>
<tr>
<td><strong>Dermatologic disease</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.90 (1.46-2.48)</td>
<td>.001</td>
<td>1.47 (1.06-2.02)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1.90 (1.46-2.48)</td>
<td>.001</td>
<td>1.47 (1.06-2.02)</td>
<td></td>
</tr>
<tr>
<td><strong>Acne</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>2.23 (1.06-4.70)</td>
<td>.03</td>
<td>1.50 (0.72-3.14)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2.23 (1.06-4.70)</td>
<td>.03</td>
<td>1.50 (0.72-3.14)</td>
<td></td>
</tr>
<tr>
<td><strong>Atopic dermatitis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3.63 (2.83-4.65)</td>
<td>.001</td>
<td>2.59 (1.96-3.45)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3.63 (2.83-4.65)</td>
<td>.001</td>
<td>2.59 (1.96-3.45)</td>
<td></td>
</tr>
<tr>
<td><strong>Other dermatitis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3.55 (2.64-4.78)</td>
<td>.001</td>
<td>2.48 (1.79-3.43)</td>
<td>.03</td>
</tr>
<tr>
<td>No</td>
<td>3.55 (2.64-4.78)</td>
<td>.001</td>
<td>2.48 (1.79-3.43)</td>
<td>.03</td>
</tr>
<tr>
<td><strong>Psoriasis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3.22 (1.57-6.62)</td>
<td>.001</td>
<td>2.23 (1.26-3.96)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3.22 (1.57-6.62)</td>
<td>.001</td>
<td>2.23 (1.26-3.96)</td>
<td></td>
</tr>
<tr>
<td><strong>Psychiatric disorder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.68 (1.06-2.69)</td>
<td>.03</td>
<td>1.25 (0.77-2.02)</td>
<td>.03</td>
</tr>
<tr>
<td>No</td>
<td>1.68 (1.06-2.69)</td>
<td>.03</td>
<td>1.25 (0.77-2.02)</td>
<td>.03</td>
</tr>
<tr>
<td><strong>Depressive Disorder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3.77 (2.26-6.30)</td>
<td>.001</td>
<td>2.62 (1.56-4.39)</td>
<td>.001</td>
</tr>
<tr>
<td>No</td>
<td>3.77 (2.26-6.30)</td>
<td>.001</td>
<td>2.62 (1.56-4.39)</td>
<td>.001</td>
</tr>
<tr>
<td><strong>Eating behavior disorder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>7.67 (5.34-11.02)</td>
<td>.001</td>
<td>6.02 (3.88-9.16)</td>
<td>.001</td>
</tr>
<tr>
<td>No</td>
<td>7.67 (5.34-11.02)</td>
<td>.001</td>
<td>6.02 (3.88-9.16)</td>
<td>.001</td>
</tr>
<tr>
<td><strong>Obsessive-compulsive disorder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3.12 (36.4)</td>
<td>.001</td>
<td>3.21 (1.57-6.62)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3.12 (36.4)</td>
<td>.001</td>
<td>3.21 (1.57-6.62)</td>
<td></td>
</tr>
<tr>
<td>Factor</td>
<td>Univariate risk ratio (95% CI)</td>
<td>$P$ value</td>
<td>Multivariate risk ratio (95% CI)</td>
<td>$P$ value</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------------------------------</td>
<td>-----------</td>
<td>----------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td><strong>Attention-deficit/hyperactivity disorder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>18/59 (30.5)</td>
<td>2.49 (1.41-4.40)</td>
<td>.003</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>304/2032 (15)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Anxiety disorder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>108/292 (37)</td>
<td>4.35 (3.30-5.74)</td>
<td>.001</td>
<td>2.30 (1.36-3.88)</td>
</tr>
<tr>
<td>No</td>
<td>214/1799 (11.9)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Plastic surgery</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>57/355 (16.1)</td>
<td>1.06 (0.78-1.45)</td>
<td>.71</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>298/355 (83.9)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

$^a$Ref: reference value.  
$^b$Not applicable.

Given the heterogeneity of the population, most of whom were women aged 18-24 years, we performed a multivariate analysis considering gender and age (Table 3). On the one hand, after multivariate analysis by gender in the group of women, students, eating disorder, and anxiety disorder remained with a significance at <.05 as diagnostic predictors of BDD (Table 3). In the male group after multivariate analysis, none of the factors analyzed showed statistical significance. On the other hand, in the multivariate analysis by age (Table 3), the following factors remained with significance at <.05 as diagnostic predictors of BDD. In the 18-24 years analysis, the diagnostic predictors were female gender and other gender, students, depressive disorder, eating disorder, and anxiety disorder. In the 25-44 years analysis, the diagnostic predictor was income level between €2000 and €2999. In the 45-64 years analysis, the diagnostic predictor was female gender, and the >64 years analysis was not performed due to the small sample size.
Table 3. Multivariate analysis by group.

<table>
<thead>
<tr>
<th></th>
<th>BDD&lt;sup&gt;a&lt;/sup&gt; prevalence, n/N (%)</th>
<th>Risk ratio (95% CI)</th>
<th>P value</th>
<th>Multivariate risk ratio (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Female gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>233/752 (30.9)</td>
<td>Ref&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worker</td>
<td>43/541 (7.9)</td>
<td>0.19 (0.14-0.27)</td>
<td>.001</td>
<td>0.46 (0.22-0.96)</td>
<td>.04</td>
</tr>
<tr>
<td>Other</td>
<td>9/128 (7)</td>
<td>0.17 (0.08-0.34)</td>
<td>.001</td>
<td>0.26 (0.09-0.72)</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Monthly income (€)&lt;sup&gt;d&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;500</td>
<td>181/617 (29.3)</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500-999</td>
<td>30/107 (28)</td>
<td>0.94 (0.59-1.48)</td>
<td>.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000-1999</td>
<td>34/255 (13.3)</td>
<td>0.37 (0.25-0.55)</td>
<td>.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000-2999</td>
<td>8/192 (4.2)</td>
<td>0.10 (0.05-0.22)</td>
<td>.001</td>
<td>0.37 (0.15-0.93)</td>
<td>.04</td>
</tr>
<tr>
<td>&gt;3000</td>
<td>91/173 (5.2)</td>
<td>0.13 (0.07-0.26)</td>
<td>.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Eating behavior disorder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>69/125 (55.2)</td>
<td>6.195 (4.23-9.07)</td>
<td>.001</td>
<td>4.03 (2.32-6.98)</td>
<td>.001</td>
</tr>
<tr>
<td>No</td>
<td>215/1296 (16.6)</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Anxiety disorder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>103/250 (41.2)</td>
<td>3.83 (2.85-5.16)</td>
<td>.001</td>
<td>2.94 (1.65-5.25)</td>
<td>.001</td>
</tr>
<tr>
<td>No</td>
<td>181/1171 (15.5)</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>18-24 years age group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>236/787 (30)</td>
<td>4.20 (2.75-6.42)</td>
<td>.001</td>
<td>3.19 (2.05-4.97)</td>
<td>.001</td>
</tr>
<tr>
<td>Male</td>
<td>27/292 (9.2)</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>4/12 (33.3)</td>
<td>4.91 (1.39-17.37)</td>
<td>.01</td>
<td>6.38 (1.72-23.67)</td>
<td>.006</td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>255/1005 (25.4)</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worker</td>
<td>8/67 (11.9)</td>
<td>0.40 (0.19-0.85)</td>
<td>.02</td>
<td>0.39 (0.17-0.86)</td>
<td>.02</td>
</tr>
<tr>
<td><strong>Dermatologic disease</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>177/640 (27.7)</td>
<td>1.53 (1.15-2.05)</td>
<td>.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>90/451 (20)</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Psychiatric disorder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>130/301 (43.2)</td>
<td>3.62 (2.70-4.86)</td>
<td>.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>137/790 (17.3)</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Depressive Disorder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>73/152 (48)</td>
<td>3.55 (2.49-5.06)</td>
<td>.001</td>
<td>1.71 (1.04-2.84)</td>
<td>.04</td>
</tr>
<tr>
<td>No</td>
<td>194/930 (20.7)</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Eating disorder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>66/107 (61.7)</td>
<td>6.27 (4.12-9.54)</td>
<td>.001</td>
<td>3.70 (2.13-6.42)</td>
<td>.001</td>
</tr>
<tr>
<td>No</td>
<td>201/984 (20.4)</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Anxiety disorder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>93/194 (47.9)</td>
<td>3.83 (2.76-5.30)</td>
<td>.001</td>
<td>2.14 (1.22-3.78)</td>
<td>.008</td>
</tr>
<tr>
<td>No</td>
<td>174/897 (19.4)</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: <sup>a</sup> BDD = Body Dysmorphic Disorder, <sup>b</sup> Ref = Reference, <sup>c</sup> Indicates a significant difference, <sup>d</sup> Monthly income is reported in €.
### Monthly income (€)

<table>
<thead>
<tr>
<th>Income Range</th>
<th>BDD prevalence, n/N (%)</th>
<th>Risk ratio (95% CI)</th>
<th>P value</th>
<th>Multivariate risk ratio (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;500</td>
<td>9/35 (25.7)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>500-999</td>
<td>3/17 (17.6)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>1000-1999</td>
<td>15/109 (13.8)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2000-2999</td>
<td>274/2.7</td>
<td>0.08 (0.02-0.40)</td>
<td>.002</td>
<td>0.12 (0.02-0.86)</td>
<td>.04</td>
</tr>
<tr>
<td>&gt;3000</td>
<td>1/43 (2.3)</td>
<td>0.07 (0.01-0.57)</td>
<td>.01</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

### 45-65 years age group

**Gender**

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

**Dermatologic disease**

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

**Other dermatitis conditions**

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

**a**BDD: body dysmorphic disorder.  
**b**Ref: reference value.  
**c**Not applicable.  
**d**€1=US $1.11.

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

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

Discussion

Principal Findings

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

Our study reports that the prevalence of BDD in adults is 15.2% in Spain, which is higher than that reported in the general population (0.5%-3.2%) in another study [1] and higher than that reported in a Spanish multicentric study in patients with acne (10.6%) [2]. Our findings may be explained by 2 factors.

First, our sample population was particularly young, with more than half of the participants being in the age group of 18-24 years. As shown in a previous study [6], the younger the age of the patient, the greater is the possibility of BDD diagnosis. Second, our data reflect more of a screening diagnosis as compared with a definitive diagnosis established with BDDQ via an interview by a health professional, which is more demanding [17].

Our study shows that the sociodemographic characteristics most associated with the diagnosis of BDD are gender (female and other), age group of 18-24 years, students, income level of less than €500/month, and a diagnosis of previous dermatologic or psychiatric disease. Regarding gender, females showed a statistically significant ($P<.001$) higher prevalence of BDD than males (284/1421, 19.9% vs 34/653, 5.2%, respectively). Previous studies have reported similar prevalence between men and women [3] or increased prevalence in women (68.5%) compared to that in men [4]. In our study, almost 67.9% (1421/2091) of the participants were women, which may suggest the need for future studies controlled by sex to clarify the differences. Regarding age, the mean age at diagnosis of BDD in the participants in our study was 23.5 years, which is higher than the mean age of 16.9 years described elsewhere [5]. This is probably because our study only included populations 18
years and older, resulting in an increase in the mean age at diagnosis. Furthermore, we found a higher prevalence of BDD among students (263/1043, 25.2%), which was also higher than 1.3%-5.8% reported in another study [1]. Again, this could be because almost half our sample was comprised of students. Moreover, this can be attributed to the influence of social media in the current age, increasing young people’s concern about their body image [1]. It would be appropriate to conduct a more specific study including this young population group.

Participants with BDD in our study were concerned about an average of 4.6 different body parts, which is in line with that reported by previous studies (5-7 body parts) [13]. The body parts that were of the most concern were body fat (248/322, 77%), thighs (191/322, 59.3%), face (166/322, 51.6%), hip (144/322, 44.7%), and skin (126/322, 39.1%). Skin (14/26, 54%) and face (8/26, 31%) were the body parts of the greatest concern in previous studies [1]. The prevalence of BDD in participants with dermatologic conditions (204/1072, 19%) in our study falls within the range (4.9%-21.1%) reported in the literature [1]. However, it is necessary to emphasize that previous studies did not distinguish between dermatologic patients per se and those undergoing aesthetic procedures [1].

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

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

Limitations and Strengths

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

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

Conclusion

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

Acknowledgments

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

Authors’ Contributions

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

https://formative.jmir.org/2024/1/e46515 JMIR Form Res 2024 | vol. 8 | e46515 | p.26
Conflicts of Interest

In the last 24 months, HBF has received lecture fees from Takeda, BIAL, laboratorios Rubio, and laboratorios Rovi. He has also been granted 3 prizes for the development of a serious videogame for treating attention-deficit/hyperactivity disorder (The Secret Trail of Moon) called as the Shibuya Prize by Takeda, first prize of the College of Psychologists of Madrid, and a prize for the Best Innovative Health Initiative within Healthstart. He is the Principal Investigator of an iPFIS research contract (accessed on August 12, 2022; IFI16/00039), Coprincipal Investigator of a MINECO research grant (RTI2018-101857-B-I00), and Principal Investigator of a research of the Sincronia project, funded by the start-up Bitsphi; recipient of a Fund for the Improvement of Postsecondary Education grant and an IDIHPISA intensification grant; involved in 2 clinical trials (Mensia Koala, Newrofeed Study; ESKETSUI2002); and a cofounder of Haglaia Solutions. He is also an employee and member of the advisory board of Ita Salud Mental (Korian).

References


Abbreviations

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

©Álvaro Loewen, Hilario Blasco-Fontecilla, Chao Li, Marcos Bella-Fernández, Belén Ruiz-Antorán. Originally published in JMIR Formative Research (https://formative.jmir.org), 04.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Original Paper

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

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

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

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

Abstract

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

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

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

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

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

https://formative.jmir.org/2024/1/e51839
KEYWORDS
alcohol reduction; alcohol; digital care; digital intervention; ethnic minority; methods; mHealth; randomised controlled trial; recruitment; retention; social media

Introduction

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

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

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

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

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

Methods

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

Design

Participants

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

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

Initially, as well as through 3 emails (days 0, 5, and 9) and (from January 15, 2022) a total of 2 SMS text messages (days 5 and 9), we had planned that at the 1-, 3-, and 6-month follow-up, all participants would also be sequentially offered opportunities to complete follow-up through phone (called twice from days 10 to 17), a mailed survey (from day 18), and a mailed postcard (from day 30). However, due to resource constraints, from November 2020 on, we only used automated emails on days 0, 5, 9, and 11 to contact participants at the 1- and 3-month follow-up; we no longer called or sent postal surveys. With the aim of improving these follow-up rates with less resource, we added attention checks, whereby individuals were asked to select a certain response option. Participants were also asked to confirm their age at 2 different points in the baseline survey to ensure they were consistent. Individuals failing either of these attention checks were screened out of the survey before randomization. The most prevalent type of fraud were “bot” responses (n=863). These were fraudulent responses similar to manual fraud, but they occurred in batches of 20-30 at a time when contact information was given in noticeably similar formats (eg, firstname123@emailaddress.com), often with American street addresses (being UK-based was an inclusion criteria of the trial). These responses seemed to be automated and were identified using the same process of address checking as above (individuals not known at the phone number provided were removed from the study). Adding a CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) to the survey eliminated this issue. A more detailed discussion on participant deception is described elsewhere [20].

Sociodemographic Characteristics

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

Analysis

Aim 1: Methods of Recruitment

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

**Aim 2: Follow-Up**

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

**Aim 3: Broader Methodological Issues**

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

**Ethical Considerations**

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

**Results**

**Sample Characteristics**

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

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Baseline (N=5602)</th>
<th>1-month follow-up (n=3685)</th>
<th>3-month follow-up (n=3574)</th>
<th>6-month follow-up (n=4458)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>3207 (57.25)</td>
<td>2046 (55.52)</td>
<td>1992 (55.74)</td>
<td>2534 (56.84)</td>
</tr>
<tr>
<td>Women</td>
<td>2365 (42.22)</td>
<td>1620 (43.96)</td>
<td>1565 (43.79)</td>
<td>1903 (42.69)</td>
</tr>
<tr>
<td>Other</td>
<td>26 (0.46)</td>
<td>16 (0.43)</td>
<td>14 (0.39)</td>
<td>17 (0.38)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>4 (0.07)</td>
<td>3 (0.08)</td>
<td>3 (0.08)</td>
<td>4 (0.09)</td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>96 (1.71)</td>
<td>68 (1.85)</td>
<td>68 (1.9)</td>
<td>83 (1.86)</td>
</tr>
<tr>
<td>Black</td>
<td>47 (0.84)</td>
<td>35 (0.95)</td>
<td>39 (1.09)</td>
<td>41 (0.92)</td>
</tr>
<tr>
<td>Chinese</td>
<td>9 (0.16)</td>
<td>9 (0.24)</td>
<td>9 (0.25)</td>
<td>9 (0.2)</td>
</tr>
<tr>
<td>White</td>
<td>5296 (94.54)</td>
<td>3474 (94.27)</td>
<td>3361 (94.04)</td>
<td>4206 (94.35)</td>
</tr>
<tr>
<td>Mixed</td>
<td>113 (2.02)</td>
<td>75 (2.03)</td>
<td>71 (1.99)</td>
<td>84 (1.88)</td>
</tr>
<tr>
<td>Other</td>
<td>21 (0.37)</td>
<td>15 (0.41)</td>
<td>15 (0.42)</td>
<td>18 (0.4)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>19 (0.34)</td>
<td>9 (0.24)</td>
<td>11 (0.31)</td>
<td>16 (0.36)</td>
</tr>
<tr>
<td>Not known</td>
<td>1 (0.02)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (0.02)</td>
</tr>
<tr>
<td><strong>Occupation, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABC1b</td>
<td>4151 (74.01)</td>
<td>2759 (74.87)</td>
<td>2688 (75.21)</td>
<td>3337 (74.85)</td>
</tr>
<tr>
<td>C2DEc</td>
<td>1451 (25.9)</td>
<td>926 (25.13)</td>
<td>886 (24.79)</td>
<td>1121 (25.15)</td>
</tr>
</tbody>
</table>

aThe data is also reported in the main trial paper [18].

bABC1: managerial, professional, and intermediate occupations.

cC2DE: skilled, semiskilled, unskilled manual, and lowest-grade worked or unemployed.
Aim 1: Recruitment Methods, Demographic Diversity, and Cost Per Participant
Most participants recruited for this trial reported seeing it advertised on social media (2483/5602, 44.32%), the NHS website (1961/5602, 35.01%), or through radio or newspapers (745/5602, 13.3%), with all other recruitment methods accounting for 7.37% (413/5602) of the sample (Table 2).

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

<table>
<thead>
<tr>
<th>Recruitment method</th>
<th>Included sample (N=5602), n (%)</th>
<th>Men, n (%)</th>
<th>Ethnic minority group, n (%)</th>
<th>Low SES, n (%)</th>
<th>Fraudulent response, n/N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untargeted social media</td>
<td>2119 (37.82)</td>
<td>650 (30.67)</td>
<td>147 (6.94)</td>
<td>507 (23.93)</td>
<td>1020/3139 (32.49)</td>
</tr>
<tr>
<td>Targeted social media</td>
<td>364 (6.5)</td>
<td>353 (96.98)</td>
<td>13 (3.57)</td>
<td>90 (24.73)</td>
<td>8/372 (2.15)</td>
</tr>
<tr>
<td>National Health Service website</td>
<td>1961 (35.01)</td>
<td>628 (32.02)</td>
<td>76 (3.88)</td>
<td>570 (29.07)</td>
<td>123/2084 (5.9)</td>
</tr>
<tr>
<td>Radio or newspapers</td>
<td>745 (13.3)</td>
<td>591 (79.33)</td>
<td>27 (3.62)</td>
<td>167 (22.42)</td>
<td>19/764 (2.49)</td>
</tr>
<tr>
<td>Word of mouth</td>
<td>142 (2.53)</td>
<td>74 (52.11)</td>
<td>9 (6.34)</td>
<td>41 (28.87)</td>
<td>11/153 (7.19)</td>
</tr>
<tr>
<td>Google</td>
<td>159 (2.84)</td>
<td>44 (27.67)</td>
<td>7 (4.4)</td>
<td>50 (31.45)</td>
<td>9/168 (5.56)</td>
</tr>
<tr>
<td>Smoke Free email</td>
<td>55 (0.98)</td>
<td>13 (23.64)</td>
<td>3 (5.45)</td>
<td>11 (20)</td>
<td>10/65 (15.38)</td>
</tr>
<tr>
<td>Health care provider or general practitioner</td>
<td>15 (0.27)</td>
<td>5 (33.33)</td>
<td>3 (20)</td>
<td>4 (26.67)</td>
<td>16/31 (51.61)</td>
</tr>
<tr>
<td>Other</td>
<td>42 (0.75)</td>
<td>7 (16.67)</td>
<td>2 (4.76)</td>
<td>11 (26.19)</td>
<td>0/42 (0)</td>
</tr>
</tbody>
</table>

The percentage of participants recruited from each method (ie, N=included sample value).

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

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

Those recruited from health care providers (15/15, 100%), Smoke Free email (51/55, 93%), and word of mouth (126/142, 88.7%) appeared to have the highest response rates and those recruited through advertisements on Google (109/159, 69%), and the NHS website (1513/1961, 77%) appeared among the lowest. Table 4 presents the follow-up rates at 1-, 3-, and 6-month follow-up.
Table 3. Total cost per participant and cost per participant who were men, of ethnic minority groups, and lower socioeconomic status (SES) by recruitment method for participants of the iDEAS (iOS Drink Less, Evaluating the Effectiveness of an Alcohol Smartphone) randomized controlled trial. A currency exchange rate of £1=US $1.26988 is applicable.

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

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

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

Aim 2: Retention During Sequential Follow-Up

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

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

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

bThe cost here is the average of 2 research staff salaries (£19.77) × research hours.

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

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

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

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

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

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

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

\(^e\) Not available.

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

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

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

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

Aim 3: Broader Methodological Insights

Retention

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

Recruitment

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

Advertisement Development

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

PPI Feedback on Advertising

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

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

**Advertising and Participant Deception**

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

**Negative Engagement With Advertising**

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

**Contamination**

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

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

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

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

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

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

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

###Discussion

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

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

Implications

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

Previous Research

The findings of this study are in line with other studies that have focused on methodological issues in remote studies and RCTs [11,13]. The recruitment strategy undertaken was informed by a previous smoking cessation trial, which recommended using a range of sources but also monitoring the success of strategies throughout to recruit a large, diverse sample [11]. We have reported on the success of each strategy here to inform the planning of future trials. An additional potential strategy that we did not use here to improve ethnic diversity in trial participation is geotargeting of social media advertisements in geographic areas with an ethnically diverse population [22]. The multistage follow-up strategy and stepped approach to incentives (eg, an additional £12 if completed within 24 hours at 6 months) undertaken throughout the IDEAS trial were also informed by previous research [11]. The need to have ongoing strategies to detect participant deception in web-based studies and trials is also supported in other studies, and other strategies recommended beyond those we used are to check participant IDs during onboarding and undertake IP address checks [13].

Limitations

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

Conclusion

Most participants in this remote RCT were recruited through advertisements on social media (2483/5602, 44.32%), the NHS website (1961/5602, 35.01%), and through radio or newspapers (745/5602, 13.3%). Most recruitment methods oversampled participants who were more advantaged, women, and White. Targeted approaches through social media and radio advertising were successful in recruiting men but less successful in appealing to a more diverse demographic in terms of ethnicity and SES. There was evidence that the sequential approach taken to 6-month follow-up was successful, with 79.58% (4458/5602) follow-up rates at 6 months. This study offers recommendations for achieving balance in methodological challenges when conducting remote RCTs. Recruitment methods should be broad and targeted to achieve sociodemographic diversity. Automated emails with substantial financial incentives can achieve excellent follow-up rates of approximately 70%, but sequential offline follow-up can further boost retention by nearly 10% overall. SMS text messages can be a low-cost, low-effort way to improve follow-up rates. An important and broader takeaway is the importance of continuously monitoring, identifying, reacting to, and documenting new methodological challenges as they appear over the course of a trial. This is necessary not only to improve individual trials but also because pooling shared experiential learning can help research teams who are planning future trials.
Acknowledgments
This study is funded by the National Institute for Health and Care Research (NIHR; Public Health Research Programme, #127651). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. MO’s salary is funded by NIHR (NIHR127651) and Medical Research Council (MRC; MR/W026430/1). CG is funded by NIHR (NIHR302923).

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

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

References
17. iDEAS methods paper protocol 1.4.docx. OSFHOME. URL: [https://osf.io/aswmt](https://osf.io/aswmt) [accessed 2023-11-30]


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

Abbreviations

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

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

Please cite as:


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

JMIR Form Res 2024;8:e51839

URL: [https://formative.jmir.org/2024/1/e51839](https://formative.jmir.org/2024/1/e51839)

doi:10.2196/51839

PMID:38180802

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

Carmen Alvarez, Subhash Aryal, Elizabeth Vrany, Maria Jose Sanchez, Rosalphie Quiles, Lia Escobar-Acosta, Felicia Hill-Briggs

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

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

Abstract

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

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

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

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

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

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

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

Introduction

Background

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

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

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

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

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

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

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

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

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

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

Procedures

Women who provided consent were found eligible for the study (based on the baseline study questionnaire), and women who agreed to enroll in the study were randomized to receive either Cuidándome or educational content from a health promotion manual designed in Mexico [31]. Randomization was stratified based on ACE score so that one group would not have more people with higher average ACEs than the other. After completion of the baseline questionnaire and randomization, participants were told when their group sessions would begin, and participants were mailed the corresponding workbook for

---

**Figure 1.** CONSORT flowchart of participant enrollment, allocation, and adherence. CONSORT: Consolidated Standards of Reporting Trials; GAD-7: Generalized Anxiety Disorder-7; PHQ-8: Patient Health Questionnaire-8.
their group assignment. Based on input from community partners and the significant use of mobile phones and apps in the study population, we did not make computer or internet access a requirement for participation. For those who were not familiar with Zoom, a brief orientation was scheduled to explain how to use Zoom. All participants were encouraged to join the group sessions when they were scheduled; if a participant could not attend the group session, a make-up session would be scheduled with the participant where the facilitator would review content from the missed week prior to the next group session. This progression was particularly important because the sessions were designed to build on each other.

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

**Intervention**

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

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

<sup>a</sup>ACE: adverse childhood experience.
<sup>b</sup>PTSD: posttraumatic stress disorder.

Comparison

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

**Group Facilitators and Intervention Fidelity**

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

**Study Questionnaire**

**Demographic Characteristics**

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

**Adverse Childhood Experiences**

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

**Primary Outcome Variables**

**Depression**

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

**Anxiety**

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

**Social Problem-Solving**

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

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

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

**Results**

**Participant Demographics, ACEs, and Retention**

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

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

\(^a\)Baseline sample.

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

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

Depression and Anxiety and Social Problem-Solving Overview

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

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

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

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

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

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

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

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

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

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

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

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

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

<table>
<thead>
<tr>
<th></th>
<th>Depression</th>
<th>Anxiety</th>
<th>Social problem-solving</th>
<th>Negative problem orientation</th>
<th>Avoidance style</th>
<th>Impulsive-careless style</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wilks A</strong></td>
<td>F test (df)</td>
<td>Wilks A</td>
<td>F test (df)</td>
<td>Wilks A</td>
<td>F test (df)</td>
<td>Wilks A</td>
</tr>
<tr>
<td>Intervention by time interaction</td>
<td>0.96</td>
<td>0.57 (3, 37)</td>
<td>0.96</td>
<td>0.48 (3, 37)</td>
<td>0.99</td>
<td>0.92</td>
</tr>
<tr>
<td>Time main effect</td>
<td>0.53</td>
<td>11.1 (3, 37)</td>
<td>0.51</td>
<td>11.9 (3, 37)</td>
<td>0.51</td>
<td>11.9 (3, 37)</td>
</tr>
<tr>
<td>Intervention main effect</td>
<td>N/A</td>
<td>0.91 (1, 39)</td>
<td>N/A</td>
<td>4.3 (1, 39)</td>
<td>N/A</td>
<td>0.84 (1, 39)</td>
</tr>
</tbody>
</table>

\(^{a}P<.001.\)
\(^{b}P<.005.\)
\(^{c}P=.02.\)
\(^{d}N/A: not applicable.\)

**Depression**

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

**Anxiety**

Based on the GAD-7, anxiety levels also decreased from baseline to post intervention and remained below baseline through 6 months for both groups (Table 3). We estimated small effect sizes (Cohen d) at each time point (postintervention (T1): d=0.30 and 0.36; medium (T3): d=0.65) effect sizes for reduced anxiety symptoms (Table 3). The reduction in anxiety symptoms over time was significant, with both groups showing a reduction in symptoms across the follow-up time points (Wilks Λ=0.51; F=11.9; P<.001; Table 5). There was also a significant difference in the reduction of symptoms between the 2 groups, where Cuidándome was demonstrated to be more effective than the comparison program for reducing anxiety symptoms (F=4.3; P<.001).

**Social Problem–Solving Styles**

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

**Negative Problem Orientation**

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

**Avoidance Style**

ASs decreased from baseline to all 3 time points (Table 4). Similar to NPO, the reduction in AS over time was significant for both groups (Wilks Λ=0.68; F=5.8; P<.005; Table 5). There was no difference in change over time between the 2 groups, and there was no significant difference in the effect of the 2 programs for reducing ASs; however, the difference in mean scores postintervention (Cuidándome: mean 2.1, SD 2.17 vs comparison: mean 4.2, SD 4.81; P=.06) and at 6-month follow-up (Cuidándome: mean 1.65, SD 2.5 vs comparison: mean 4.0, SD 4.80; P=.06) approached significance, where Cuidándome participants reported lower AS at these time points (Table 4). We estimated medium effect sizes (T1: d=0.62, T3: d=0.61) for the intervention.

**Impulsive-Careless Style**

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

**Discussion**

**Principal Findings**

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

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

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

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

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

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

Limitations

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

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

Conclusions

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

Acknowledgments

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

Data Availability

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

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT e-HEALTH (V.1.6.1) checklist.
[PDF File (Adobe PDF File), 410 KB - formative_v8i1e52969_app1.pdf ]

References


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


Abbreviations

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

Roopan Gill¹,²,³, BSc, MD; Gina Ogilvie²,⁴, MD, PHD; Wendy V Norman²,⁴, MD, CCFP, DTM&H, MHSC; Brian Fitzsimmons³,³, MD; Ciana Maher², BA; Regina Renner²,³, MD, MPH

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

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

Abstract

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

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

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

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

Conclusions: This study demonstrates that communities and individuals are important collaborators in developing a mobile innovation that facilitates access to high-quality patient-centered abortion care. Through the cocreation process, a digital platform such as myPostCare highlighted a gap in abortion care in Canada, particularly around follow-up support after a procedural abortion.
KEYWORDS

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

Introduction

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

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

The New England Journal of Medicine published a special report on Telehealth in the United States, highlighting its utility and future. In 2016, Kaiser Permanente of Northern California reported that its virtual (email, telephone, and video) communications had exceeded in-person visits [11]. Similarly, research supports the safe and effective use of telehealth for the provision of medication abortion care globally [12-14].

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

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

Methods

Participants

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

Study Design

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

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

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

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

Data Analysis

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

Ethical Considerations

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

Results

Participant Characteristics

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

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

Quantitative

Change in Contraceptive Method

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

System Usability Scale

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

Satisfaction

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

![Graphical representation of the satisfaction results in percentage.](https://example.com/satisfaction_graph.png)

Qualitative

Overview

Qualitative analysis of the interviews was completed using thematic analysis including both inductive and deductive themes. Nine key themes were identified and are listed in Textbox 1.
Ease of Use
Overall, there was unanimous agreement that myPostCare was easy to use with an organized and easy-to-navigate design. One participant highlighted the following:

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

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

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

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

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

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

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

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

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

The participants unanimously stated that myPostCare provided them the support that was needed at the right time. It was helpful for the resource to provide support over time and that it allowed them to navigate various aspects of their postcare journey. A few participants shared the following:

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

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

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

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

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

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

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

In addition, a participant commented that the website had credible information, which helped to answer questions that she would have seen her physician about and, therefore, kept her out of the office or emergency room. When further asked if the website helped her understand when to seek hospital care, she mentioned that it was very clear. She specifically found the disclaimer useful in preventing women from misunderstanding the website as a substitute for clinical care:

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

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

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

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

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

**Frequency of Use**

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

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

**Time Spent on Website**

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

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

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

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

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

**Suggestions for Improvement of myPostCare**

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

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

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

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

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

**Recommend to a Friend**

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

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

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

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

**Privacy and Security**

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

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

**Design Features**

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

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

Yes, I was really happy I signed up for it and I was getting those e-mails weekly. I was able to access it, once again, read about different perspectives. I think Overall, I don’t know if this is weird to say, but it was very calming. Approachable in a sense. It doesn’t hurt your eyes to stay on the website for a while. I really liked the colours. The layout was easy. [participant 6]

Other participants noted that the site was structurally thought through and that the design was relatable to them. One participant highlighted the following:

I thought it was clearly thought out and structurally too. The language is nicely worded and was very unbiased. [participant 2]

It was very nice and pleasant, the pictures were very cool. I liked the ranges of things that were on there, the whole thing about meditating and then also just needing actual straight up information was really helpful too. [participant 2]

Very soothing colours. The nice blues and then greys, yes. [participant 2]

I thought it was very easy to use. I felt that the information was well laid out with the menu sidebar. The images were quite big and spaced out so had to scroll quite a bit and not get through a lot of information. [participant 3]

There was no harsh colours. There was no in-your-face type of things that popped into the website. I liked there were no advertisements. I think the peaceful colourings, the “click this if you feel called to.” It’s nice to have that sense of well-being with positivity on a sensitive topic, it was well crafted. [participant 3]

**Overall Impressions of myPostCare**

Overall, the participants were satisfied with myPostCare. They felt supported by the resource. There were very strong sentiments that this went above and beyond what they had expected. One participant stated that she was surprised to have such a good experience with a website, as she had never had such an experience. Some of the participants used the site with their partners and appreciated the section that was specifically for partners. One participant stated that the emails and website helped to keep her out of the emergency department, as it highlighted the normalcy of postprocedural recovery. In general, all participants felt that it was a great experience to have this resource, and many participants expressed that this resource should be available as long as women need abortions:

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

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

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

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

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

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

Web-Based Secure Analytics

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

<table>
<thead>
<tr>
<th>Web-based analytics data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique visitors to the website, n (%)</td>
</tr>
<tr>
<td>Average time spent on the website by visitors</td>
</tr>
<tr>
<td>Longest visit on the website</td>
</tr>
<tr>
<td>Total number of page views, n</td>
</tr>
<tr>
<td>Average daily page views, n</td>
</tr>
<tr>
<td>Participants who are mobile users, n (%)</td>
</tr>
<tr>
<td>Participants who are desktop visitors, n (%)</td>
</tr>
<tr>
<td>Participants who visited the Emotional Well-Being page, n (%)</td>
</tr>
<tr>
<td>Participants who visited the Contraception Explorer page, n (%)</td>
</tr>
<tr>
<td>Participants who visit the Postprocedure Care page, n (%)</td>
</tr>
<tr>
<td>Participants who visit the Sexual Health page, n (%)</td>
</tr>
<tr>
<td>Top 3 pages on the website</td>
</tr>
<tr>
<td>Most popular contraceptives visited from the contraception tool in page views, n</td>
</tr>
<tr>
<td>Hormonal IUD&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Sterilization</td>
</tr>
<tr>
<td>Copper IUD</td>
</tr>
<tr>
<td>Vaginal ring</td>
</tr>
<tr>
<td>Fertility awareness</td>
</tr>
<tr>
<td>Patch</td>
</tr>
<tr>
<td>Abstinence</td>
</tr>
<tr>
<td>Depo shot</td>
</tr>
<tr>
<td>Female condom</td>
</tr>
<tr>
<td>Male condom</td>
</tr>
<tr>
<td>Withdrawal</td>
</tr>
<tr>
<td>Visits to given feelings (good, okay, and not so good) from the Emotional Well-Being tool&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Okay</td>
</tr>
<tr>
<td>Good</td>
</tr>
<tr>
<td>Not so good</td>
</tr>
<tr>
<td>Visits to given emotion from the Emotional Well-Being tool, n&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Grief</td>
</tr>
<tr>
<td>Relief</td>
</tr>
<tr>
<td>Supported</td>
</tr>
<tr>
<td>Sadness</td>
</tr>
<tr>
<td>Guilt</td>
</tr>
<tr>
<td>Regret</td>
</tr>
<tr>
<td>Shame</td>
</tr>
</tbody>
</table>

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

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

Email Notifications Analytics

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

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

Discussion

Principal Findings

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

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

We adopted a few theoretical frameworks, all of which use a comprehensive participatory approach to developing eHealth technologies. This was similarly performed by Gilbert et al [38] in the development of Get Checked Online, a web-based sexually transmitted infection testing resource. More specifically, integrating the Technology Acceptance Model and Theory of Reasoned action with the human-centered design methodology, we used a holistic approach to developing myPostCare. According to the Technology Acceptance Model, perceived ease of use and perceived usefulness of a system are the 2 predominant indicators of system adoption [27,32].

Participants in our study were accustomed to using some form of technology, either mobile phones or computers; did not require acquisition of new skills; and were keen on developing a technology-based tool to support follow-up care after an abortion. Importantly, myPostCare will not eliminate structural barriers to comprehensive abortion care, and although it may not directly affect health behavior and decision-making, it may assist in making the delivery of abortion care more efficient, convenient, patient centered, and accessible.

myPostCare is a unique addition to the literature because of its methodology and outcome. There is evidence to support eHealth technologies to improve health care; however, currently there is limited research on mobile interventions specifically to address postabortion care, although there are various interventions for contraception use. A randomized trial in Cambodia demonstrated that involving women in the design and testing of a mobile intervention to support postabortion contraception led to more women in the intervention group reporting use of effective contraception at 4 months; specifically, the use of long-acting contraceptives was higher in the intervention group at 4 and 12 months after the procedure [40].

Previous feasibility trials focused on usability and acceptability have highlighted the importance of conducting a pilot study first, which can then assist with the design of a larger randomized trial to measure effectiveness [41]. Finally, similar to studies on the development and testing of contraception tools, the integration of evaluation in real-time clinical care is essential to ascertain the barriers and challenges to implementation in the future [22].

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

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

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

Conclusions

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

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

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

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

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

Multimedia Appendix 1
Demographic data.

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


27. Gill R, Ogilvie G, Norman WV, Fitzsimmons B, Maher C, Renner R. Feasibility and acceptability of a mobile technology intervention to support postabortion care (the FACTS study phase II) after surgical abortion: user-centered design. JMIR Form Res 2024 | vol. 8 | e46284 | p.70https://formative.jmir.org/2024/1/e46284


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


Abbreviations

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

Edited by A Mavragani; submitted 05.02.23; peer-reviewed by D Levine, B Payne, H McMahon; comments to author 15.09.23; revised version received 03.11.23; accepted 05.11.23; published 09.01.24.

Please cite as:

Feasibility and Acceptability of a Mobile Technology Intervention to Support Postabortion Care After Surgical Abortion (the FACTS Study Phase 3): Mixed Methods Prospective Pilot Study
JMIR Form Res 2024;8:e46284
URL: https://formative.jmir.org/2024/1/e46284
doi:10.2196/46284
PMID:38194252

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

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

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

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

Abstract

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

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

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

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

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

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

KEYWORDS

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

Introduction

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

Diabetes and Mobile Apps

Even if technical solutions to support diabetes self-management, such as smartphone apps, have become more common, they are used by a minority of patients [6]. It is not clear how many people use diabetes apps in Sweden, but in Australia, only 8% of people reportedly use diabetes mobile apps [6], despite almost unlimited availability with thousands of apps on the market. Reasons for people with diabetes not to use apps can be unawareness of their existence, technical literacy barriers, no need (the disease is not that bad or self-management is sufficient anyway), no recommendation from a health care professional, the resulting increased accountability for one's own behaviors, or the time-consuming nature of some apps [7,8]. However, studies have shown that people with type 2 diabetes want to use smartphone apps, to reduce not only the practical burden but also the cognitive and emotional burden of diabetes self-management [9]. Studies also have shown that patients want to have more contact with their nurse or physician through digital media than is the case today [9,10]. The most effective app-based technical solutions, in terms of the potential to reduce hemoglobin A\textsubscript{1c} (HbA\textsubscript{1c}), are interactive and include components such as patient-generated health data, individualized feedback, 2-way communication, and tailored education [11,12]. These components are in line with the 2 behavior change techniques of “feedback on behavior” and “self-monitoring of behavior” that are associated with better glycemic control [13].

Figure 1. How the diabetes app is intended to improve glycemic control.

Sukaribit Smartphone App

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

Aim

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

Methods

Research Design

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

Ethical Considerations

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

Progression Criteria

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

Participants and Procedures

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

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

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

All questionnaires were administered using the Uppsala University Psychosocial Care Program (U-CARE) Portal (the portal). The participants answered the questionnaires at the time of randomization and 8 weeks later (a delayed response of a maximum of 18 days was allowed). Randomization occurred in the portal (see the following paragraph), was totally automated, and occurred in blocks of 6 immediately after the completion of baseline questionnaires.

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

Preliminary Effect Sizes

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

HbA1c was analyzed from a home testing blood test at an accredited laboratory. The blood glucose measurements were recorded by the participants in their own diary of choice and reported in the portal as an outcome. The intervention group could use the app to record their measurements. Physical activity was measured as steps via pedometers on the participants’ smartphones, and the last 7 days were reported in the portal.
Participants also reported the number of occasions per week over the last month they had exercised more than 30 minutes for fitness purposes. Diabetes self-management was measured using the Diabetes Self-Management Questionnaire (DSMQ) [20,21]. The DSMQ has 16 items divided into 4 subscales, namely (1) glucose management, (2) dietary control, (3) physical activity, and (4) health care use, with a maximum score of 64. A higher score indicates higher frequency of diabetes self-care behaviors. DSMQ has shown good psychometric properties in several contexts [21]. The Diabetes Distress Scale (DDS) was used to measure diabetes distress [22]. The DDS has 17 items divided into 4 subscales, namely (1) emotional burden, (2) physician-related distress, (3) regimen-related distress, and (4) diabetes-related interpersonal distress, with a maximum score of 102. A higher score indicates more distress. DDS has shown good psychometric properties in several contexts [22].

The Intervention

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

Data Analysis

The collected data on recruitment and intervention use were compared with the prespecified progression criteria to decide if they matched. Qualitative data were analyzed with quantitative content analysis [23]. Data from both the interviews and open-ended questions were analyzed together. Within and between-group effect sizes (Cohen d) were calculated for HbA1c and self-reported outcomes, dividing the mean differences with pooled SDs, with the aim of being the basis for statistical power and sample size calculations for a future study. The between-group effect sizes used the pooled baseline SDs as recommended by Morris [24]. A value of d>0.8 is classified as a large effect size, d=0.5 is classified as a medium effect size, and d=0.2 is classified as a small effect size according to Cohen [25]. Preliminary inference statistics were also performed utilizing linear regression analysis with the posttreatment value as the outcome and group allocation, baseline values, sex, and age included as covariates. The adjusted estimate can be interpreted as the adjusted mean difference for the treatment group when compared with the control group (the reference). P<.05 was considered significant.

Results

Participant Characteristics

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

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total sample</th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants randomized</td>
<td>—</td>
<td>28 (48)</td>
<td>31 (53)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>17 (29)</td>
<td>11 (39)</td>
<td>6 (19)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>61.1 (10)</td>
<td>60.2 (12)</td>
<td>61.8 (9)</td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>15 (25)</td>
<td>7 (25)</td>
<td>8 (26)</td>
</tr>
<tr>
<td>Cohabiting/married</td>
<td>41 (70)</td>
<td>20 (71)</td>
<td>21 (68)</td>
</tr>
<tr>
<td>Living alone but have a steady partner</td>
<td>3 (5)</td>
<td>1 (4)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Country of birth, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>54 (92)</td>
<td>25 (89)</td>
<td>29 (94)</td>
</tr>
<tr>
<td>Outside Sweden</td>
<td>5 (9)</td>
<td>3 (11)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Education level, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>8 (14)</td>
<td>1 (4)</td>
<td>7 (23)</td>
</tr>
<tr>
<td>Secondary</td>
<td>17 (29)</td>
<td>10 (36)</td>
<td>7 (23)</td>
</tr>
<tr>
<td>University (≤3 years)</td>
<td>17 (29)</td>
<td>7 (25)</td>
<td>10 (32)</td>
</tr>
<tr>
<td>University (&gt;3 years)</td>
<td>17 (29)</td>
<td>10 (36)</td>
<td>7 (23)</td>
</tr>
<tr>
<td>Employment status, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Unemployed</td>
<td>2 (3)</td>
<td>0</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Retired</td>
<td>32 (54)</td>
<td>14 (50)</td>
<td>18 (58)</td>
</tr>
<tr>
<td>Employed (any status)</td>
<td>25 (42)</td>
<td>14 (50)</td>
<td>11 (36)</td>
</tr>
<tr>
<td>Employed full time</td>
<td>23 (39)</td>
<td>12 (43)</td>
<td>11 (36)</td>
</tr>
<tr>
<td>Employed part time</td>
<td>2 (7)</td>
<td>2 (7)</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Exercise intensity, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mostly sedentary</td>
<td>13 (22)</td>
<td>7 (25)</td>
<td>6 (19)</td>
</tr>
<tr>
<td>Lightly active</td>
<td>35 (59)</td>
<td>17 (60)</td>
<td>18 (58)</td>
</tr>
<tr>
<td>Moderately active</td>
<td>8 (14)</td>
<td>3 (10)</td>
<td>5 (16)</td>
</tr>
<tr>
<td>Very active</td>
<td>3 (5)</td>
<td>1 (4)</td>
<td>2 (6)</td>
</tr>
<tr>
<td>Days per week with ≥30 minutes of physical activity, mean (SD)</td>
<td>2.5 (2)</td>
<td>2.3 (2)</td>
<td>2.7 (2)</td>
</tr>
<tr>
<td>Steps per day in the last week, mean (SD)</td>
<td>4966 (3862)</td>
<td>4798 (3164)</td>
<td>5094.5 (4371)</td>
</tr>
<tr>
<td>Current smoker (Yes), n (%)</td>
<td>6 (10)</td>
<td>2 (7)</td>
<td>4 (13)</td>
</tr>
<tr>
<td>Medical history, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>39 (66)</td>
<td>16 (57)</td>
<td>23 (74)</td>
</tr>
<tr>
<td>Dyslipidemia</td>
<td>27 (46)</td>
<td>11 (39)</td>
<td>16 (52)</td>
</tr>
<tr>
<td>Stroke</td>
<td>1 (2)</td>
<td>0</td>
<td>1 (3)</td>
</tr>
<tr>
<td>History of mental illness</td>
<td>8 (14)</td>
<td>3 (11)</td>
<td>5 (16)</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>4 (7)</td>
<td>2 (7)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Other cardiovascular disease</td>
<td>8 (14)</td>
<td>6 (21)</td>
<td>2 (7)</td>
</tr>
</tbody>
</table>
### Table 2. Summary of the progression criteria with goals and study values.

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

Feasibility of Study Procedures

Table 2 summarizes the progression criteria fulfilment. There were 182 people that reported interest in participating in the study; of this group, the majority (176/182, 96.7%) registered their interest on a web page. That met progression criterion 1 (n≥60) by a good margin. Of the 182 people interested, 133 were reached and assessed for eligibility. However, a considerable proportion of the participants who registered their interest were ineligible or unable to be included in the study; hence, progression criterion 2 (50% inclusion rate) was not met. The main reason for exclusion at this stage was a BMI <25 kg/m² (n=26). In total, 55 people were excluded. Of those eligible, 19 people never logged into the portal even after being reminded. Finally, 59 (76%) of the 78 eligible participants were randomized in the study (treatment: n=28; control: n=31). For the complete recruitment flow, see Figure 2.
Progression criterion 3, at least 75% complete data at follow-up, was met regarding questionnaire data (81%). However, when also considering HbA1c tests, the completeness was 64%; thus, the criterion was not met. At baseline, 5 HbA1c test results were missing. At follow-up, 11 randomized participants did not complete their questionnaires, and there were 18 missing tests (Table 3). For baseline HbA1c, 41 manual reminders were sent in total; for the follow-up HbA1c, 63 manual reminders were sent. See Table 3 for details.
Table 3. Feasibility data (n=59).

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

\textsuperscript{a}HbA\textsubscript{1c}: hemoglobin A\textsubscript{1c}.

\textsuperscript{b}Missing tests + defective tests: n=18.

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

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

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

Client Satisfaction and the Physician’s Evaluation
A summary of the interviews is presented in Table 4. The findings show that smartphones with the iOS operating system were the most commonly used among the responding participants (15/20, 75%). Concerning the overall quality of the app, a majority of the participants reported the app was of fair quality, with only a few of their individual needs having been met. The 4 technical issues reported to the developer mainly concerned the iOS version of the app. The physician had technical problems but thought the contact was rewarding when it worked. She also experienced varying activity from the participants (Table 4).
Table 4. Participants’ (n=20) answers from the telephone interviews and open-ended questions after completing the intervention, as well as the physician’s (n=1) evaluation.

<table>
<thead>
<tr>
<th>Questions and categories</th>
<th>Participants, n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expectations for the app and study</strong></td>
<td></td>
</tr>
<tr>
<td>Want to have contact with a physician or health care professional (feedback)</td>
<td>8</td>
</tr>
<tr>
<td>Interest in diabetes apps</td>
<td>6</td>
</tr>
<tr>
<td>Contribute to research</td>
<td>4</td>
</tr>
<tr>
<td>Thinking that a diabetes app is part of the future for diabetes care</td>
<td>3</td>
</tr>
<tr>
<td>Help with more motivation to perform self-care</td>
<td>2</td>
</tr>
<tr>
<td>Want more knowledge</td>
<td>3</td>
</tr>
<tr>
<td>Ability to collect everything in the same place (though it is not working)/facilitate everyday life</td>
<td>2</td>
</tr>
<tr>
<td><strong>Thoughts about the app</strong></td>
<td></td>
</tr>
<tr>
<td>Technical problems</td>
<td>12</td>
</tr>
<tr>
<td>Difficult to add their medicine in the list</td>
<td>8</td>
</tr>
<tr>
<td>Technical problems when sending messages/values to the physician</td>
<td>3</td>
</tr>
<tr>
<td>Thought the app was difficult</td>
<td>6</td>
</tr>
<tr>
<td>Did not like the appearance of the app</td>
<td>2</td>
</tr>
<tr>
<td>Easy to navigate</td>
<td>8</td>
</tr>
<tr>
<td>Simple but functioning</td>
<td>2</td>
</tr>
<tr>
<td>Easier to manage more frequent blood glucose monitoring</td>
<td>3</td>
</tr>
<tr>
<td>Easy access to and communication with health care staff</td>
<td>3</td>
</tr>
<tr>
<td>Possible to get feedback on test results from physician</td>
<td>3</td>
</tr>
<tr>
<td>Increased motivation for self-care/increased awareness</td>
<td>3</td>
</tr>
<tr>
<td>Interesting to see how blood glucose is affected by food</td>
<td>1</td>
</tr>
<tr>
<td>Possibility to log data/follow data over time</td>
<td>3</td>
</tr>
<tr>
<td>Good support from the developer</td>
<td>2</td>
</tr>
<tr>
<td><strong>Contact with the physician</strong></td>
<td></td>
</tr>
<tr>
<td>Good and relevant replies</td>
<td>9</td>
</tr>
<tr>
<td>Good contact and fast communication</td>
<td>3</td>
</tr>
<tr>
<td>Some sort of miscommunication due to technical issues and maybe a lack of personal knowledge</td>
<td>1</td>
</tr>
<tr>
<td>No/very little communication with the physician</td>
<td>4</td>
</tr>
<tr>
<td><strong>Desired improvements</strong></td>
<td></td>
</tr>
<tr>
<td>Wish for an easier app</td>
<td>4</td>
</tr>
<tr>
<td>Improved design</td>
<td>1</td>
</tr>
<tr>
<td>Faster and more communication with caregiver</td>
<td>2</td>
</tr>
<tr>
<td>Direct communication between the app and blood glucose meter</td>
<td>5</td>
</tr>
<tr>
<td>Linked to other health applications</td>
<td>3</td>
</tr>
<tr>
<td>See old values and a graph function (to be able to learn)</td>
<td>4</td>
</tr>
<tr>
<td>Notifications when receiving message overview/table/graph</td>
<td>2</td>
</tr>
<tr>
<td>Wanted bigger text or a computer version</td>
<td>2</td>
</tr>
<tr>
<td>Information/news about diabetes in the app</td>
<td>3</td>
</tr>
<tr>
<td>Be able to send photos</td>
<td>1</td>
</tr>
<tr>
<td>Be able to log physical activities</td>
<td>1</td>
</tr>
<tr>
<td>Be able to set goals</td>
<td>2</td>
</tr>
</tbody>
</table>
**Questions and categories**

Be able to add notes to values
No function was missing

<table>
<thead>
<tr>
<th>Participants, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

**Overall impression**

The application did not improve self-care.
The app improved self-care.
The participants were positive about the concept and think the app should continue to be developed.

<table>
<thead>
<tr>
<th>Participants, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

**The physician’s evaluation**

Lots of technical problems (messages, medicine list)
The contact and work were fun when the app worked.
Disadvantage not being their attending physician
The app as a good complement to diabetes care; could consider using it with her own patients
Varying participation of the participants; some very active but others never replied
Room for many improvements
Part of the future

<table>
<thead>
<tr>
<th>Participants, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

**Effect Sizes of Outcome Measures**

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

<table>
<thead>
<tr>
<th>Outcome measures</th>
<th>Treatment group</th>
<th>Control group</th>
<th>Between group (post), (d)</th>
<th>Linear regression analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline, mean (SD)</td>
<td>Post, mean (SD)</td>
<td>n</td>
<td>Baseline, mean (SD)</td>
</tr>
<tr>
<td><strong>HbA(_1c) (mmol/mol)</strong></td>
<td>50.1 (6.91)</td>
<td>49.3 (7.15)</td>
<td>20</td>
<td>-0.11</td>
</tr>
<tr>
<td><strong>Blood glucose tests per day</strong></td>
<td>1.27 (1.62)</td>
<td>1.23 (1.14)</td>
<td>20</td>
<td>-0.03</td>
</tr>
<tr>
<td><strong>Physical activity per week(^c)</strong></td>
<td>3.25 (1.8)</td>
<td>3.15 (1.8)</td>
<td>20</td>
<td>-0.06</td>
</tr>
<tr>
<td><strong>Steps per day</strong></td>
<td>4761 (3099)</td>
<td>5407 (3117)</td>
<td>18</td>
<td>0.21</td>
</tr>
<tr>
<td><strong>EQ-5D-VAS (1-100)</strong></td>
<td>55.1 (22.8)</td>
<td>57.8 (25.5)</td>
<td>18</td>
<td>0.11</td>
</tr>
<tr>
<td><strong>Total DDS(^d)</strong></td>
<td>2.45 (0.83)</td>
<td>2.35 (0.76)</td>
<td>20</td>
<td>-0.12</td>
</tr>
<tr>
<td><strong>Total DSMQ(^e)</strong></td>
<td>6.62 (1.34)</td>
<td>6.59 (1.53)</td>
<td>19</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

\(^a\)The posttreatment between-group effect size was adjusted for baseline values.

\(^b\)The difference between groups after treatment was adjusted for age, sex, and baseline values of the respective measure. The control group is the reference.

\(^c\)Number of times, in the last month, the participant performed exercise for more than 30 minutes.

\(^d\)DDS: Diabetes Distress Scale.

\(^e\)DSMQ: Diabetes Self-Management Questionnaire.

Discussion

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

Feasibility of Recruitment and Data Collection

Considering the recruitment of participants, the number of responses to the advertisements met and even exceeded that specified in criterion 1. However, the proportion of people who could not be reached or were ineligible to participate due to criterion 2 was slightly higher than ideal. Some alterations to the exclusion criteria were already made during the recruitment phase of the study (ie, including people with an HbA\(_1c\) <50 mmol/mol or age >65 years). This could potentially have resulted in the inclusion of participants with relatively well-managed diabetes (HbA\(_1c\) <50 mmol/mol) and participants with less technological experience (age >65 years) The remaining exclusion criterion of a BMI <25 kg/m\(^2\) resulted in the largest number of exclusions. This was thought to exclude participants who would not likely benefit from the intervention. Another way to facilitate the recruitment process could be to add inclusion or exclusion questions on the study-specific website to better be able to reach the right target group.

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

Although data collection from self-reported questionnaires worked well, meaning that progression criterion 3 was partially met, the collection of HbA\(_1c\) data through home testing kits did not work well. The first problem was the long administration time. The mean time from sending the kits out until the results were received by the project team was 16 days, with the main delay appearing to be at the participant’s home. There was also a large amount of missing data due to both defective tests and missing tests, even though several manual reminders were sent. Hence, there is a need to make the collection of HbA\(_1c\) data more reliable and efficient. Previous studies have also reported difficulties using these test kits [26]. Better or additional instructions, more telephone reminders, another test kit brand, or another lab are things to consider. Most likely, the biggest advantage can be gained by improving the participant handling of the test and posting. Other ways to handle this could be to...
conduct this kind of trial within the health care system so that the blood test is managed by health care professionals and not by the participant. Another thing that we could have done differently is to not have HbA1c as an inclusion criterion. In a full-scale trial, the participant’s glycemic control at the start of the intervention could be a minor issue, since it is the effect of the intervention (the difference) that is measured.

Feasibility and Acceptability of the Intervention

Not all eligible participants who signed up for the trial started the intervention (progression criterion 4). We do not know the reasons for this; possibly, they just regretted the enrollment. When or if conducting a larger study, the possibility that not all who are accepting of study participation will actually join the study needs to be considered. A few of the participants in the qualitative evaluation thought that the app improved self-care, but the majority did not think so. Many participants appeared to have been less active than anticipated, especially based on the amount of diagnostic data and messages sent to the physician. This was progression criterion 5, which was not met. Some participants described technical issues that interfered with the use of the app (eg, lack of access to pedometer data in the app as well as difficulties logging medications and viewing summary features). These problems could most often be related to the iOS version. This could have had an impact on user motivation leading to less activity. Participants also suggested improvements in the message function and added features when logging data (eg, in the calendar function, graph) in order to make the app more user-friendly. For the app to be beneficial, it is important that it is used. Previous studies [12] have shown that unsatisfied users will be less active and therefore will not benefit from using this kind of app. Multistep tasks, difficult system navigation, limited functionality, and limited interaction are generally the most common and important usability problems.

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

Effect Sizes

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

Clinical Significance

The results of this study demonstrate the importance of conducting a feasibility trial in order to avoid unnecessary financial as well as study burden for those involved. In order to proceed with a larger clinical trial, a number of problems both in study design and the intervention, as described in the previous sections, need to be addressed. The next step then is to perform a sufficiently powered RCT. If the results are favorable, this will be the first step toward clinical evidence for the intervention, and a new digital treatment helping people to better manage their type 2 diabetes may be available shortly [28]. The participants who signed up for this study were particularly interested in mobile apps; therefore, the results from this study are applicable for patients with type 2 diabetes who want a digital aid. The app could complement standard care and possibly increase empowerment and self-care management. The main advantage of this app is that it enables a new and, maybe, faster way for communication between the person with diabetes and the diabetes nurse or general practitioner. This app, along with other available apps, could be suitable for health care now as well as in future, more digital health care [29].

Limitations

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

Strengths

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

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

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

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

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

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

Conflicts of Interest
None declared.

References


Abbreviations

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

Sophie Rutter\(^1\), PhD; Sally Sanger\(^1\), PhD; Andrew D Madden\(^1\), PhD; Sukaina Ehdeed\(^1\), PhD; Catherine Stones\(^2\), PhD

\(^1\)Information School, University of Sheffield, Sheffield, United Kingdom
\(^2\)School of Design, University of Leeds, Leeds, United Kingdom

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

Abstract

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

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

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

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

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

\(\text{(JMIR Form Res 2024;8:e47308) doi:}\ 10.2196/47308\)

Keywords

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

https://formative.jmir.org/2024/1/e47308
Introduction

Hand Hygiene in Workplaces

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

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

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

Smart Sanitizers

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

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

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

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

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

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

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

**Methods**

**Overview**

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

**Recruitment**

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

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

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

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

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

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

Ethical Considerations
This study received ethics approval (038337) from the University of Sheffield Research Ethics Committee on February 16, 2021. All study participants received an information sheet about the project, and they were given opportunities to ask questions and advised that they could withdraw with no negative
Results

RQ1: What Actions Could Be Taken From HH Data?

Overview

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

Manage Hygiene Resources and Workflows

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

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

Find Operating HH Facilities

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

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

Communicate the (High) Standard of Organizational Hygiene

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

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

Improve Own HH Practice

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

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

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

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

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

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

Promote an Organizational Hygiene Policy

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

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

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

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

This would also apply when new staff join the organization.

Enforce Organizational HH Policy

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

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

**Target HH Training According to Needs**
An analysis of HH data could also help to identify who needs training and what their training needs are:

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

**Execute More Effective Interventions**
Organizations and researchers could evaluate the effectiveness of interventions in real time and adapt them according to the results:

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

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

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

**No Action**
Whether any action would be taken was also discussed. HH was often thought to be a personal matter and the responsibility of the individual. Several participants stated that they would not say anything or take any action if they knew their colleagues had inadequate HH practices:

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

**RQ2: For the Actions Identified, What, if Any, Are the Concerns of Office Workers?**
Office workers are concerned that (1) intentions and messages could be misinterpreted and that data could be used for (2) bullying and coercion, (3) unintended purposes, (4) inaccurate representation of HH practices, and (5) incomplete representation of HH practices. Next, we have discussed which actions raise the concerns.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Incomplete Representation of HH Practices

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

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

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

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

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

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

If it went like comparing groups within an organisation, then how do you know you’re comparing like with like? [P5]

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

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

When Identity Is Protected

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

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

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

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

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

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

Situations Where HH Is Considered Important

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

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

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

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

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

When Events Considered Private Are Not Recorded

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

Some people might think that it’s a bit of an invasion of privacy, being monitored in the toilet as well. Is there nowhere safe? Is there nowhere that I can just not be monitored? [P4]

When Data Collection Can Be Customized

It was thought that office workers would be more likely to accept monitoring if they can customize the system and control what data are collected and how they are presented:

I think perhaps like with the alerts, perhaps [they] could have the option to turn that on [recommendations, encouragement, advice] if you so desired, but it shouldn’t be a requirement. [P5]

The system should allow users to correct any errors in data collection and add explanations, so that managers do not unfairly target individuals:

So that you’ve got the chance to correct yourself if you need to, like, you see I would be going back to my computer and I would expect then to have a message on my computer that says “You’ve walked past a hand sanitiser and you didn’t use it. What was the reason?” [P4]

When Data Are Used for a Beneficial Purpose

Participants felt that data need to be collected for a purpose. The purpose needs to be justified, and the data should be retained only for as long as necessary. Furthermore, systems should be used to support individuals rather than punish them:

If there was a real, proper reason that they were collecting it for, then they could collect it for the relevant time period. So if there was some sort of disease outbreak and it lasted six months, then collect it for six months...it has to be justifiable...it’s not right to just collect it and hold that data. [P5]

It depends on what people perceive is the overall intention of whoever’s putting this policy in place. If it’s used--., if the perception is it’s used to beat people up about hygiene because it’s going to lead to a poor sick record or more transmissible covid than that is a different intention to “Well, I’m really bothered about how sore your fingers are becoming with all your hand washing.” [P12]

When Uses of the Collected Data Are Clearly Communicated

Participants recommended that the uses of the data need to be transparent and clearly communicated including how the data will be used and reported, who has access to what data, where and for how long the data are stored, and whether it is possible to opt out of data collection:

Why they’re doing it, who’s going to hold the data, who’s going to see it, how’s it going to be reported, who will it be shared with, yeah, where will the data go, how long will they hold the data for, and can I opt out, how do I opt out. [P2]

Discussion

Principal Findings

Our previous study found that office workers thought HH data should only be collected if they can be acted upon [22]. In this study, office workers were able to identify several actions that could usefully be informed by HH data (RQ1). These included using the data to manage hygiene resources and workflows, find operating sanitizers, communicate (high) organizational standards of hygiene, improve workers’ own practice, promote and enforce an organization’s hygiene policy, target the training according to needs, execute more effective interventions, and identify the causes of outbreaks. However, hygiene is mostly considered as a private matter, and it is possible that no action would be taken in practice. Furthermore, office workers expressed concerns (RQ2) that the data could be used to bully, to coerce and for unintended purposes. Moreover, the data could be misinterpreted, inaccurate, and an incomplete representation of hygiene practices. Office workers suggested that they would be more likely to accept monitoring for the identified uses (RQ3) when their privacy is respected, they have some control over how their data are collected, and how their data will be used is clearly communicated. Monitoring is also more likely to be accepted in situations where hygiene is considered important and there is a clear beneficial purpose for data collection.

HH Is (Mostly) a Personal Matter for Office Workers

Although the findings of this study suggest that facility managers, health and safety officers, departmental managers, building occupants and visitors, hygiene resource suppliers, researchers, and those interested in public health would find HH data useful, HH was thought to be a personal matter [25]. Several office workers reported that they would not take any action if they found that their colleagues’ HH practices were inadequate. For all the uses of HH data, there was a strong
preference for personal data to be anonymized or not collected at all. However, acceptance of HH monitoring is dynamic and dependent on the situation and the context within which the data are used. The findings of our study suggest that it is much more likely to be accepted during disease outbreaks, in certain locations (eg, entrance to buildings), and in sectors (eg, health and food) where HH is important to the ethos and culture of the organization (ie, will also influence office workers). Although none of the office workers (0/18, 0%) thought it acceptable to enforce HH by restricting access to areas (by syncing HH data with door entry system data), it is possible that under extreme circumstances and in certain locations, this could be acceptable.

Need for a Shared Understanding of “Good” Office HH

For those concerned about catching infectious diseases in the workplace [10], HH data could be used to reassure visitors to a building about the high standards of hygiene within the building and to promote and enforce organizational HH policies. In health care settings, smart sanitizers are already used to audit compliance and enforce the health sector’s policy of sanitizing hands before, during, and after patient care. Overall, 3 factors are likely to make smart sanitizers more acceptable in health care settings. First, they are used to enforce an HH policy that is considered important in professional practice [18]. Second, smart sanitizers can collect data that measure the compliance with policies such as the 5 moments for hygiene [26] that can be measured using the sanitizer supplied by the organization. Third, at least at a basic level, adherence to this policy can be monitored using room and sanitizer sensors, without the need for personal data collection.

For smart systems to be adopted in offices, it would help if there were an agreed-upon understanding of what is good hygiene practice, for example, how often and where (eg, entrances to buildings and exits from toilets) hygiene should be performed. Smart systems could then be used to reassure and promote HH in offices, if the policy can be complied with using office resources and without the need to collect personal data.

Office Workers Want Insights From All Their HH Practices

Health care workers may wish to track their HH practices around patients, as good HH is part of their professional identity [18], and having access to their personal HH data could help health care workers improve their HH practice [27,28]. Given that office workers are likely to be overestimating their HH practices [21], self-tracking could be beneficial. However, office workers in this study did not link self-tracking of HH with professional expectations and standards; rather, they were interested in gaining insights into their overall HH practices including in all locations (office, home, and when they are out of the office and across all facilities (sanitizer, soap, wipes, etc). No smart system (as yet) can automatically detect HH with such detail. This would only be possible if office workers were prepared to input data manually, and this would require considerable motivation.

Office Workers Share Health Care Workers’ Concerns

Many of the concerns that office workers expressed are similar to those of health care workers. Similar to health care workers [19,20], most office workers are concerned that personal data could be misused to punish or bully employees, and use of HH data should be clearly and transparently communicated. It would be advisable to consult with office workers early in the system design process to engage them, explain what purposes the organization intends for the data, and identify what purposes they feel are acceptable and useful.

Another concern shared with health care workers is whether smart HH systems can accurately represent HH practices, because, first, systems can be gamed and deliberately misused, and second, systems may incorrectly interpret events as missed HH opportunities because they are not registering the wide context within which the event did or did not occur [18-20]. Given the shift to more hybrid and flexible working [29], it may also be necessary to analyze HH data alongside work patterns. More generally, studies of workplace tracking have found that systems that enable employees to customize and control what data are collected are more likely to be accepted [30].

Office workers were also concerned that the data could be used for purposes other than what it was intended for. This is understandable given that misuse of technologies is widely reported in the media; for example, AirTags designed to track property have been used to stalk individuals [31].

Useful for Health Researchers

Using sensors to remotely collect HH data resolves some of the challenges for health researchers who need to evaluate hygiene interventions. Researchers may evaluate interventions by observing HH practices, but the presence of an observer may change the behavior of the person being observed, particularly because HH is a social norm. Instead, researchers may use proxy measures such as changes in soap consumption. However, manually collecting soap consumption data from organizations is time consuming, and sensors can help in saving time [32].

Although it is thought likely that poor hygiene could contribute to disease transmission in workplaces, little data are available to support (or oppose) this point [9]. Connecting smart HH system data with other health data (such as data relating to employee absence) could help researchers understand the relationship.

Data collected from digital technologies (such as mobile phones, social media networks, and search engines) have been used to communicate public health messages and monitor and control outbreaks [33]. Smart HH systems could usefully be added to the arsenal of digital data sources that have been used to support health authorities’ response to COVID-19 and any future pandemics.

When Personal Data Are or Are Not Needed

To a large extent, how well a smart HH system is accepted depends on whether personal data are collected. We next consider what data smart HH systems can collect and what is needed for the identified actions (Table 3).
For some uses of HH data, there is no need for personal data to be collected. For resource and workflow planning and to find operating sanitizers, data about soap levels and soap consumption including date, time, and location of use can be collected from sensors in smart systems without the need to collect any personal data. Similarly, smart systems can give immediate feedback to users to improve their HH technique without collecting any personal data. It is also possible to communicate the overall standard of HH in an organization without collecting personal data. Smart system data combined with data from sensors that track movement in and out of spaces can be used to identify the extent to which all employees are practicing HH and whether they use HH facilities as they move around the building (eg, at the lift and after using the toilet). The data collected could also indicate overall compliance with organizational policy and be used to identify the overall education and training needs of the organizations. The same system could also give feedback to individuals at the point where they are using the facility.

Personal data are required for several of the uses of HH data identified in this study, particularly the uses where individuals are pinpointed such as identification of individual practices and training needs and enforcement of organizational policy. Sensors that track individuals (installed on staff ID cards, apps, or other personal devices) would be necessary to capture each person’s use of a hand sanitizer (date, time, technique, and location of use) and to send reminders. However, these data can be anonymized and aggregated to identify the uses of HH facilities by different groups (not individuals) within the organization. Good practice would be to offer a manual override that would allow employees to correct any system errors. For smart systems to fully represent a person’s HH practice, it would be necessary to allow users to manually input their use of any and all hygiene facilities such as wipes and their own sanitizer gel. To capture a person’s contextualized use of HH facility (including what the person had been doing and where they had been), the system would need to connect to other personal data such as calendars and mobile phones.

Whether personal data are needed to evaluate interventions will depend on the nature of the intervention and what needs to be evaluated. Identifying the causes of outbreaks will likely also require the collection of other personal data (eg, who is ill).

Limitations

Through this interview study, a wide range of applications for HH data collected via smart systems has been identified. However, this is an exploratory study; further investigation is needed to determine whether office workers would use smart sanitizers for the identified purposes. Such studies could build on these findings to further investigate the implementation and adoption of smart sanitizers, with trials in offices.

Conclusions

Smart sanitizers could, feasibly, make a contribution to the improvement of hygiene practices in offices [21], but for smart systems to be accepted, any data collected would need to be actionable [22]. This study contributes to knowledge by identifying the many potential uses for hygiene data collected from smart systems. As smart HH systems have not yet been introduced into offices, identification of constructive uses for data is important for their design and implementation.

Although smart sanitizers are widely deployed in clinical settings, health care workers recognize that HH is an important part of their professional practice [18,19]. Given that office workers consider HH to be a mostly personal matter, it seems less likely that they will want to adopt smart sanitizers. When there are disease outbreaks, office workers may consent to the sharing of personal data and the monitoring of their own and their colleagues’ HH. At other times, smart sanitizers may need

---

**Table 3.** Data collected by smart hand hygiene (HH) systems and whether these include personal data (from sensors that track individuals, manual input, and personal devices) or do not include personal data (from sensors in dispensers of sanitizer, soap, etc and sensors [including cameras] that track anonymized movement and location).

<table>
<thead>
<tr>
<th>Data to be collected</th>
<th>No personal data</th>
<th>Personal data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensors in dispensers of sanitizer, soap, etc</td>
<td>Sensors that track individuals</td>
</tr>
<tr>
<td>Soap levels and soap consumption including date, time, and location of use</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Correct HH technique</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>HH events or nonevents</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Date, time, duration, and location of a person’s use of a hand sanitizer</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Date, time, duration, and location of a person’s use of a hygiene facility</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Contextualized use of HH facility (including what the person had been doing and where they had been)</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
to be restricted to uses that do not require any personal data collection. Should organizations wish to implement smart sanitizers in offices, it would be advisable to consult widely with the staff and to develop systems that are customizable and personalizable. It should also be noted that office workers may find it more useful to have insights from all their HH practices, but these data cannot (yet) be automatically collected from smart systems.

In contrast to health care workers, as yet, there is no widely accepted HH policy for office workers. Future studies could usefully investigate what office workers would consider to be an appropriately high standard of hygiene and how often and where hygiene should be performed. A better understanding of what would be effective and acceptable HH policies in nonclinical settings would help to clarify how smart systems can be used and hence inform their design. Importantly, it could help workplaces adopt infection prevention and control strategies that are necessary for current and future infectious disease outbreaks [1].

Acknowledgments
This study was funded by Innovate UK (grant 8154) and Economic and Social Research Council’s (ESRC’s) Impact Acceleration Account (project 172894). This Innovate UK–funded study is part of a large project conducted in collaboration with Savortex, a manufacturer of hygiene technology. However, the authors of this paper have no financial or nonfinancial interests in this study.

Data Availability
This study involved human research participant data and could contain potentially identifying information. The data that support the findings of this study are available upon request from the corresponding author.

Authors' Contributions
SR and CS were involved in conceptualization. SS, SR, and ADM contributed to the methodology. SS, SR, ADM, and SE were involved in formal analysis. SS and SE were involved in investigation. SR was involved in writing the original draft. SS, ADM, SE, and CS were involved in reviewing and editing the paper. SR contributed to project administration. SR and CS were involved in funding acquisition. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest
None declared.

References


29. Matei A. ‘I was just really scared’: Apple AirTags lead to stalking complaints. Guardian. 2022. URL: https://www.theguardian.com/technology/2020/feb/10/apple-airtags-stalking-complaints-technology [accessed 2023-12-04]


Abbreviations

AHHMS: automated hand hygiene monitoring systems
HH: hand hygiene
RQ: research question

©Sophie Rutter, Sally Sanger, Andrew D Madden, Sukaina Ehdeed, Catherine Stones. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Population-Level Portal-Based Anxiety and Depression Screening Perspectives in HIV Care Clinicians: Qualitative Study Using the Consolidated Framework for Implementation Research

Daniela Zimmer¹, MPH, MSW; Erin M Staab¹, MPH; Jessica P Ridgway¹, MD; Jessica Schmitt¹, MSW; Melissa Franco¹, MPH; Scott J Hunter¹,², PhD; Darnell Motley¹, PhD; Neda Laiteerapong¹, MS, MD

¹Section of General Internal Medicine, University of Chicago, Chicago, IL, United States
²Western Institutional Review Board- Copernicus Group, Princeton, NJ, United States

Corresponding Author:
Daniela Zimmer, MPH, MSW
Section of General Internal Medicine
University of Chicago
5841 S Maryland Ave
Chicago, IL, 60637
United States
Phone: 1 312 702 8847
Email: dazimmer@bsd.uchicago.edu

Abstract

Background: Depression and anxiety are common among people with HIV and are associated with inadequate viral suppression, disease progression, and increased mortality. However, depression and anxiety are underdiagnosed and undertreated in people with HIV owing to inadequate visit time and personnel availability. Conducting population-level depression and anxiety screening via the patient portal is a promising intervention that has not been studied in HIV care settings.

Objective: We aimed to explore facilitators of and barriers to implementing population-level portal-based depression and anxiety screening for people with HIV.

Methods: We conducted semistructured hour-long qualitative interviews based on the Consolidated Framework for Implementation Research with clinicians at an HIV clinic.

Results: A total of 10 clinicians participated in interviews. In total, 10 facilitators and 7 barriers were identified across 5 Consolidated Framework for Implementation Research domains. Facilitators included advantages of systematic screening outside clinic visits; the expectation that assessment frequency could be tailored to patient needs; evidence from the literature and previous experience in other settings; respect for patient privacy; empowering patients and facilitating communication about mental health; compatibility with clinic culture, workflows, and systems; staff beliefs about the importance of mental health screening and benefits for HIV care; engaging all clinic staff and leveraging their strengths; and clear planning and communication with staff. Barriers included difficulty in ensuring prompt response to suicidal ideation; patient access, experience, and comfort using the portal; limited availability of mental health services; variations in how providers use the electronic health record and communicate with patients; limited capacity to address mental health concerns during HIV visits; staff knowledge and self-efficacy regarding the management of mental health conditions; and the impersonal approach to a sensitive topic.

Conclusions: We proposed 13 strategies for implementing population-level portal-based screening for people with HIV. Before implementation, clinics can conduct local assessments of clinicians and clinic staff; engage clinicians and clinic staff with various roles and expertise to support the implementation; highlight advantages, relevance, and evidence for population-level portal-based mental health screening; make screening frequency adaptable based on patient history and symptoms; use user-centered design methods to refine results that are displayed and communicated in the electronic health record; make screening tools available for patients to use on demand in the portal; and create protocols for positive depression and anxiety screeners, including those indicating imminent risk. During implementation, clinics should communicate with clinicians and clinic staff and provide training on protocols; provide technical support and demonstrations for patients on how to use the portal; use multiple screening methods for broad reach; use patient-centered communication in portal messages; provide clinical decision support tools, training, and mentorship to help clinicians manage mental health concerns; and implement integrated behavioral health and increase mental health referral partnerships.
KEYWORDS
HIV; patient portal; clinic staff perspectives; depression and anxiety screening; implementation

Introduction

Barriers to Depression and Anxiety Diagnosis

Depression and anxiety are common mental health conditions among people with HIV, with a prevalence of 20% to 45% [1-10]. People with HIV experiencing symptoms of depression or anxiety are more likely to miss appointments and have lower medication adherence, higher HIV viral loads, and higher mortality rates than those without depression or anxiety [1,2,5,10,11]. However, depression and anxiety are often underdiagnosed and undertreated in people with HIV, particularly among African Americans and Hispanics, because of the perceived stigma of mental health disorders, racial discrimination, HIV-related discrimination, and medical mistrust [1,3,5,12,13]. In the HIV Cost and Service Utilization survey of people with HIV identified as experiencing depression, only 45% had a formal depression diagnosis in their medical chart [14].

For people living with chronic conditions, such as HIV, specialty care clinics often serve as their primary source of health care [3,4]. Given the frequency of visits people with HIV have with their HIV care team, establishing mental health screening in HIV clinics is a key opportunity to address depression and anxiety underdiagnosis in people with HIV [3,4]. Patients and physicians have noted that depression screening in clinics is helpful in identifying, assessing, and treating depression [3]. However, competing demands and priorities during appointments, a lack of staff to complete assessments, and a shortage of resources to offer patients after diagnosis discourage clinicians from screening and treating depression [1,3,6,7,9,15,16].

Novel Mental Health Screening

A novel strategy to increase depression and anxiety screening in people with HIV is to perform screening at the population level using the patient portal. In recent years, health care systems have increased the adoption of electronic patient portals, and patients have increasingly used portals to facilitate their health care [8,9,17-20]. Studies in primary care settings have found that depression screening rates increased significantly when clinics adopted portal-based screening [3,7,18]. Notably, a population-level portal-based depression screening intervention, in which patients were invited to complete a depression screener regardless of having a scheduled appointment, also increased depression screening and diagnosis rates [21]. This population-level portal-based approach identified more patients with moderate to severe symptoms than screening during clinic appointments [22]. Moreover, portal-based screening increases the likelihood of discussing depression diagnosis and treatment during an appointment [7,19,20,22].

Population-level portal-based screening has been shown to be promising in primary care settings but has not been examined in HIV care settings. Guided by the Consolidated Framework for Implementation Research (CFIR), we conducted and analyzed qualitative interviews with clinicians at an urban HIV clinic [23]. This study explored clinicians’ perspectives on facilitators of and barriers to implementing population-level portal-based depression and anxiety screening for people with HIV. The objective of this formative study was to use the identified facilitators and barriers to develop implementation recommendations for HIV clinics.

Methods

Study Design

We completed a qualitative study to inform the design and implementation of population-level portal-based depression and anxiety screening at an HIV clinic. This paper reports the results of interviews conducted with clinicians and clinical staff.

Study Setting

The study was conducted in the Ryan White HIV Care Clinic at an academic medical center on the South Side of Chicago, the main provider of HIV care services for Chicago South Side residents. The South Side of Chicago is one of the communities most impacted by the HIV epidemic in the United States [24]. The clinic provides care for >630 people with HIV, most of whom are African Americans and publicly insured. Currently, staffed with 15 physicians, 6 fellows, a nurse practitioner, 2 pharmacists, and 2 licensed social workers, the clinic also provides mental health services.

In November 2020, the HIV clinic adopted a protocol for conducting depression and anxiety assessments during in-person clinic visits. Medical assistants were asked to complete the 2-item Patient Health Questionnaire (PHQ) and the 2-item Generalized Anxiety Disorder (GAD) scale with patients due for annual screening, as indicated by health maintenance topics and best practice advisories in the electronic health record (EHR) [25,26]. Scores of ≥3 were reflexed into the PHQ-9 and GAD-7, respectively. Medical assistants were also asked to complete the PHQ-9 and GAD-7 with patients with a history of depression or anxiety, respectively, who were due for ongoing symptom monitoring or surveillance. Physicians and advanced practice nurses were alerted via a critical, noninterruptive best practice advisory to scores of ≥3.

Concurrent with this study, a population-level portal-based depression screening intervention was tested in the primary care clinic at the academic medical center. Patients were invited to complete depression screening using the patient portal regardless of having a scheduled appointment [21]. The clinic saw an increase in screening and identification of depression [21]. These advances in screening in the primary care clinic at the institution motivated us to gauge the interest in and feasibility of integrating population-level portal-based screening in the HIV clinic.
Study Participants
All clinicians at the HIV clinic were eligible for study enrollment, including physicians, advanced practice nurses, pharmacists, nurses, and social workers. HIV clinicians were informed about the study at a clinic meeting and by email, and if interested in participating, they were instructed to contact the study project manager. The participants verbally consented before each interview.

Data Collection
Semistructured interviews were conducted one-on-one with each participant from January to April 2021, after the newly adopted in-clinic screening protocol was implemented. Demographic information was collected via electronic surveys in the REDCap (Research Electronic Data Capture; Vanderbilt University). Interviews were conducted over Zoom (Zoom Video Communications) and lasted for approximately 60 minutes. The interview questions were created by the research team using the CFIR interview guide tool for all 5 CFIR domains. The full interview guide is available in Multimedia Appendix 1. The following are example questions by domain:

- Innovation characteristics: “Do you think assessing anxiety and depression using the patient portal will be effective? Why or why not?”
- Inner setting: “What is the general level of receptivity in the clinic to using the patient portal?”
- Outer setting: “Do you think measuring anxiety and depression using the patient portal will meet the needs of the patients served by your clinic? Why or why not?”
- Process: “Who are other key influential individuals to get on board with assessing anxiety and depressive symptoms using the portal?”

Data Analysis
Descriptive analysis was used to summarize the characteristics of the participants interviewed. Initially, the original CFIR domains and constructs from the codebook were used for our interview analysis. During the initial coding phase, research team members identified additional themes and subthemes to be added to the interview analysis. Once consensus was achieved on the codebook for our analysis, 2 independent coders analyzed each interview transcript, and coding discrepancies were discussed until a consensus was reached. Analysis of coded transcripts was performed in the web-based software Dedoose (version 9.0.17; SocioCultural Research Consultants, LLC). A total of 2 research team members independently reviewed the coded excerpts to find common themes within each domain, identified each as a facilitator or a barrier, and discussed them to consensus. On the basis of these facilitators and barriers, the 2 research team members proposed implementation strategies and presented these strategies to the entire study team for validation and refinement.

Ethical Considerations
The study was reviewed and approved by the University of Chicago Biological Sciences Division Institutional Review Board (20-1313). The research team obtained oral consent from participants before the beginning of the interviews. Interview audio was recorded, and transcripts were deidentified before qualitative coding. The data were accessible to the research team only. Participants were given a US $40 e-gift card for interview completion.

Results
Participant Characteristics
Interviews with HIV clinicians continued until the team agreed that data saturation was met, as indicated by the lack of new themes emerging in the interviews. In total, 10 HIV clinicians completed the interviews. The participants ranged in age from 31 to 64 years. Most participants were identified as White (8/10, 80%) or male (6/10, 60%). As shown in Table 1, 70% (7/10) worked as physicians at the HIV clinic, and the remaining staff included a social worker (1/10, 10%), a nurse (1/10, 10%), and a pharmacist (1/10, 10%).
Table 1. Demographic information of interview participants.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Participants (N=10), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (y)</strong></td>
<td></td>
</tr>
<tr>
<td>30-39</td>
<td>5 (50)</td>
</tr>
<tr>
<td>40-49</td>
<td>2 (20)</td>
</tr>
<tr>
<td>50-59</td>
<td>2 (20)</td>
</tr>
<tr>
<td>60-69</td>
<td>1 (10)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>6 (60)</td>
</tr>
<tr>
<td>Female</td>
<td>4 (40)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>1 (10)</td>
</tr>
<tr>
<td>White</td>
<td>8 (80)</td>
</tr>
<tr>
<td><strong>Clinical role</strong></td>
<td></td>
</tr>
<tr>
<td>Physician</td>
<td>7 (70)</td>
</tr>
<tr>
<td>Social worker</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Nurse</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Pharmacist</td>
<td>1 (10)</td>
</tr>
<tr>
<td><strong>Caring for people with HIV (y)</strong></td>
<td></td>
</tr>
<tr>
<td>1-5</td>
<td>3 (30)</td>
</tr>
<tr>
<td>6-10</td>
<td>2 (20)</td>
</tr>
<tr>
<td>11-15</td>
<td>2 (20)</td>
</tr>
<tr>
<td>≥15</td>
<td>3 (30)</td>
</tr>
<tr>
<td><strong>Clinical experience (y)</strong></td>
<td></td>
</tr>
<tr>
<td>1-5</td>
<td>5 (50)</td>
</tr>
<tr>
<td>6-10</td>
<td>2 (20)</td>
</tr>
<tr>
<td>≥15</td>
<td>3 (30)</td>
</tr>
</tbody>
</table>

**Current Depression and Anxiety Screening Practice**

When asked about their current mental health screening practices, most participants mentioned informally screening patients by asking how they were feeling or if the patient was experiencing any thoughts of self-harm or suicidal ideation. As 1 participant said, “Within the review of systems [during patient intake], I often will ask if any depression and anxiety-type symptoms [were experienced] recently, but there is no standard way I approach every patient” (Participant 6, physician). Participants reported that patients expressing depression or anxiety would typically be screened with the PHQ and the GAD questionnaire. A few participants spoke of the newly implemented in-clinic mental health screening procedure put in effect before the interviews were conducted. However, they stated that the screening protocol was not regularly followed during the clinic visits.

Similarly, the participants mentioned that initiating mental health services relied on patients requesting services or bringing up suicidal ideation or self-harm. The clinic relies on HIV-trained clinical social workers to connect patients with mental health resources based on the individual’s insurance. One participant explained, “I’ll have our social worker call them and set them up with a resource and have her—obviously she’s more trained in that than I am, I believe—and have her assess them and provide appropriate resources” (Participant 9, physician). Before the COVID-19 pandemic, patients were regularly introduced to social workers during in-clinic appointments. The participants emphasized that these in-person interactions were valuable, as they built trust between the patient, the social worker, and the physicians to increase intervention uptake.

**Perceptions of Population-Level Portal-Based Screening**

**Overview**

Facilitators of and barriers to population-level portal-based depression and anxiety screening were identified within the 5 CFIR domains (Tables 2 and 3).
Table 2. Facilitators to population-level portal-based depression and anxiety screening from qualitative interviews with HIV clinicians.

<table>
<thead>
<tr>
<th>Domain and facilitators</th>
<th>Facilitator quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Innovation characteristics</strong></td>
<td></td>
</tr>
<tr>
<td>Advantages of systematic screening outside clinic visits</td>
<td>“I like the idea that it feels like it would make it a little less stressful on the clinic visit because it’s done independently of that. And then it might come up during the visit, but at least the initial screening and questions doesn’t get added to that clinic workflow. So from that standpoint, it’s a little bit relieving. Because we do want to do this, but you don’t want to have anyone having too much work or too many things on their plate. So from that standpoint, it feels like a good method to go about it.” (Participant 2, physician)</td>
</tr>
<tr>
<td>Expectation that assessment frequency could be tailored to patient needs</td>
<td>“I think if we did it this way, then we would have the information at the beginning of a visit, and could then walk into the visit knowing this. And maybe even have some additional background from our social worker, if they’ve reached out to them in the meantime, between the time they filled this out and got these results, and then the time we see it for an appointment.” (Participant 8, physician)</td>
</tr>
<tr>
<td><strong>Outer setting</strong></td>
<td></td>
</tr>
<tr>
<td>Greater respect for patient privacy</td>
<td>“I kind of think that’s where you get the most honest answers, in the patient’s environment. In clinic, the patient’s mental status is, I’m ready for clinic. So they have that person put on, their clinic person. And unless something is like really, really outstanding, they’re not forthcoming with their information, right?” (Participant 4, nurse)</td>
</tr>
<tr>
<td>Normalizing mental health screening</td>
<td>“…what I’m thinking is I really liked the way it is being approached as making it a routine part of HIV [care]...just destigmatizing and routinizing those questions for people. I think once that becomes routine as part of your, whatever, yearly check-in, I think that’s helpful.” (Participant 3, social worker)</td>
</tr>
<tr>
<td><strong>Inner setting</strong></td>
<td></td>
</tr>
<tr>
<td>Compatible with clinic culture, workflows, and systems</td>
<td>“Overall positive feelings towards it. I like the idea that it feels like it would make it a little less stressful on the clinic visit because it’s done independently of that. And then it might come up during the visit, but at least the initial screening and questions doesn’t get added to that clinic workflow. So from that standpoint, it’s a little bit relieving. Because we do want to do this, but you don’t want to have anyone having too much work or too many things on their plate.” (Participant 2, physician)</td>
</tr>
<tr>
<td>Protocol for addressing positive screening results</td>
<td>“Would be nice to have a pathway that’s somewhat predetermined. So, it’s like, ‘Okay, we identified this patient has this. We’re not sure if they’re going to be able to see a mental health provider because it might take 2 months to get in... But in the meantime, this is the plan. This is our protocol for what we should do. These are first-line medications. This is the plan from our social work standpoint of how we’re going to follow up with them.’ So, things like that. That would make it easier once we do identify the need to take some of the guesswork out of what the next steps are.” (Participant 8, physician)</td>
</tr>
<tr>
<td><strong>Characteristics of individuals</strong></td>
<td></td>
</tr>
<tr>
<td>Participant beliefs about the importance of mental health screening and benefits HIV care</td>
<td>“Some of it could be a little bit more work within the appointment if you’re then talking about some of these issues and how they affect their other medical care, but I think it would be time likely worth spent and gratifying, and probably maybe more time spent in the front would help decrease time later needing to if it were something that could be addressed and then would improve compliance, that would be very meaningful and worth discussion.” (Participant 6, physician)</td>
</tr>
<tr>
<td>Participant interest in evidence-based practices and desire to learn from prior implementation</td>
<td>“I think there’s a strong need to do it in general.... I haven’t read all the literature on it.... Most likely this needs to happen. We need to screen people. And then the question is just, ‘What’s the best way to screen?’ And looking at everything and talking through the pros and cons, it feels like this would probably be a good way to do it.” (Participant 2, physician)</td>
</tr>
<tr>
<td><strong>Process</strong></td>
<td></td>
</tr>
<tr>
<td>Team-based approach that leverages strengths of all clinic staff</td>
<td>“Our patient population is a bit delicate, which is why we have different levels to our team approach, because what patients wouldn’t share with their doctor they will share with me, because they easily identify with me. So they accept it on an extended family member kind of like basis. So their level of trust is greater. And we use that. It’s very effective.” (Participant 4, nurse)</td>
</tr>
<tr>
<td>Clear planning and communication with staff</td>
<td>“In general having something that’s standardized is good. Having something that doesn’t totally disrupt the workflow in clinic. So using the patient portal is excellent. And having a really clear plan for what the follow-up is for the patient. I think those are the really important things. And if those are well communicated to the clinic, to the section before-hand..... We have our Monday meetings at noon, something like that.... So that way everyone’s comfortable. I would be comfortable going forward with something like this, but making sure that everyone’s on the same page.” (Participant 2, physician)</td>
</tr>
<tr>
<td>System that empowers patients to communicate about their mental health</td>
<td>“But most of our patient population is a very secretive population. So I believe being able to have something on their own terms...[The social workers could] be like, ‘Hey, if you ever feel A, B, and C.... Hey just answer these questions. I get an alert and I will respond or someone will respond in a reasonable timeframe.’ Yeah. So if patients have the information that you can use MyChart to let us know if something is going on, I think that would be more successful than just screening patients as they check-in in clinic.” (Participant 4, nurse)</td>
</tr>
</tbody>
</table>
### Table 3. Barriers to population-level portal-based depression and anxiety screening from qualitative interviews with HIV clinicians.

<table>
<thead>
<tr>
<th>Domain and barriers</th>
<th>Barrier quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Innovation characteristics</strong></td>
<td>“I’d want to make sure that with this screener that we’re assessing for suicidal or homicidal ideation and that somehow that gets like flagged to be address immediately because when you’re in the clinical setting, you can address it immediately. But over the portal, I worry that it might just like sit there, and then what happens if someone is actively suicidal and they fill this out and nobody addresses it.” (Participant 8, physician)</td>
</tr>
<tr>
<td>Difficulty of ensuring prompt response to those in imminent risk to themselves or others</td>
<td>“And the reality, with my patient population, there has been issues with just accessing MyChart for a variety of patient problems, if you will. Lack of technical skills, lack of just having no laptop or any way to do that, or just feel comfortable with that kind of thing.” (Participant 5, physician)</td>
</tr>
<tr>
<td>Limited patient access, experience, and comfort using the portal</td>
<td>“Our barrier is the resource pool that we have to select from...we’re extremely limited...if someone is not a threat to themselves or someone else, however they’re battling their issues that are too much for them to really handle, where do we refer our patients to? And the destinations are booked out. And I personally believe that time is a factor when we’re dealing with depression and anxiety.” (Participant 4, nurse)</td>
</tr>
<tr>
<td>Limited availability of mental health services</td>
<td>“I think a lot of the physicians I work with don’t even check their in-basket, answer My Chart messages...I mean, I’ve been using it and I do like it... But I think a lot of the people I work with... They trained in a different time, none of this was around then. A lot of them give out their cell phone numbers to their patients and that’s how they end up communicating.” (Participant 10, physician)</td>
</tr>
<tr>
<td>Clinician variation in the use of electronic health records</td>
<td>“We have a list of our own priorities that we need to address every visit. I think sometimes I’m like, ‘Well, they have a primary care physician. That’s the appropriate person that should assess and counsel, and hopefully they’re doing that.’ I kind of rely on that. And probably we’re often not as good at realizing when someone is in some kind of mental health distress. We see someone and they might seem like they’re doing okay and I’m like, ‘I don’t need to ask them how their mental health is.’ But obviously under the surface could be a very different story.” (Participant 9, physician)</td>
</tr>
<tr>
<td>Limited capacity to address mental health concerns during HIV visits</td>
<td>“I think that a lot of the actionable information or the action that I’ll most likely take will eventually fall on the [infectious disease] clinic social worker [based] on my previous behavior.... I’m not likely, to be honest, to start any medication. I just don’t feel well-versed enough or practiced enough to really prescribe pharmacologic interventions. So usually the interventions I would take are to refer them to their provider or have our social worker kind of provide resources in some way. I don’t feel equipped to provide nonpharmacologic interventions related to anxiety/depression or pharmacologic.” (Participant 9, physician)</td>
</tr>
<tr>
<td><strong>Characteristics of individuals</strong></td>
<td>“I think that a lot of the physicians I work with don’t even check their in-basket, answer My Chart messages...I mean, I’ve been using it and I do like it... But I think a lot of the people I work with... They trained in a different time, none of this was around then. A lot of them give out their cell phone numbers to their patients and that’s how they end up communicating.” (Participant 10, physician)</td>
</tr>
<tr>
<td>Participant concerns about limited knowledge about mental health treatments</td>
<td>“I think a lot of the actionable information or the action that I’ll most likely take will eventually fall on the [infectious disease] clinic social worker [based] on my previous behavior.... I’m not likely, to be honest, to start any medication. I just don’t feel well-versed enough or practiced enough to really prescribe pharmacologic interventions. So usually the interventions I would take are to refer them to their provider or have our social worker kind of provide resources in some way. I don’t feel equipped to provide nonpharmacologic interventions related to anxiety/depression or pharmacologic.” (Participant 9, physician)</td>
</tr>
<tr>
<td><strong>Process</strong></td>
<td>“As you know, it takes patient buy-in to be able to feel like you’re not...exploiting people almost.... You’re getting into mental health and that sometimes can be a touchy subject to do in an impersonal manner, I would imagine, through something like an email or a text or MyChart.... I can envision certain patients not really warming up to it, just because it is impersonal and you’re just filling out.... I’m sure the patients, if they have the idea that this is totally for their upcoming visit and we just want to make sure we’re being complete and we want to take care of you, if there’s any concerns in the realm of depression/anxiety, we’d like to be able to address them appropriately.” (Participant 5, physician)</td>
</tr>
</tbody>
</table>

### CFIR Domain 1: Innovative Characteristics

Codes within the innovation characteristics domain focused on the attributes of population-level portal-based screening. The participants spoke about its relative advantage, adaptability, complexity, evidence base, and design quality.

#### Facilitator: Advantages of Systematic Screening Outside Clinic Visits

The participants thought that population-level portal-based mental health screening would help make screening more consistent without imposing additional work or disrupting the clinic workflow. With screening completed ahead of time via the patient portal, the participants felt that they would be better prepared to address these concerns during the visit.

#### Facilitator: Expectation That Assessment Frequency Could be Tailored to Patient Needs

Participants generally thought that sending assessments via the portal every 6 months or once a year would be ideal; however, they mentioned that they would defer to the evidence on screening frequency best practices. The ability to send patients mental health screeners at a custom interval appealed to participants as it would keep them aware of the mental health concerns that arose. The participants also expressed interest in tailoring the screening frequency based on symptom severity.

#### Barrier: Difficulty Ensuring Prompt Response to Those in Imminent Risk to Themselves or Others

A common concern raised during the interviews was the complexity of responding to patients who indicated suicidal risk, self-harm, or homicidal ideation on the screener.
Participants were worried that patients with immediate mental health needs would not receive timely interventions if screening were performed via the portal.

**CFIR Domain 2: Outer Setting**

Codes within the outer setting domain focused on external factors that might affect the implementation of population-level portal-based screening, particularly the needs and resources of the patient population served by the clinic.

**Facilitator: Greater Respect for Patient Privacy**

Participants spoke about how patients seen at the HIV clinic value privacy. They thought that using the portal might increase screening uptake and encourage honest responses by allowing patients the flexibility to complete screening in environments where they are most comfortable.

**Facilitator: Normalizing Mental Health Screening**

Participants recognized the stigma associated with mental illness for some patients. Implementing a routine depression and anxiety screening process was seen as an approach to destigmatize mental health assessments. In addition, participants believed that consistent depression and anxiety screening would frame mental health as part of patients’ general health care, compared with the sporadic mental health assessments in current practice.

**Barrier: Limited Patient Access, Experience, and Comfort Using the Portal**

The participants did not know if patients were familiar enough with the portal to complete the assessments electronically. The participants reported that several of their patients did not know how to access their laboratory work via the portal. Therefore, they were not confident that the patients could complete screeners via the portal without assistance or training. The participants were also concerned about usability issues regarding the small text and reading levels associated with using the portal.

**Barrier: Limited Availability of Mental Health Services**

Participants emphasized that accessible mental health referral pathways and resources were needed before the clinic could implement population-level portal-based screening. Otherwise, patients would be diagnosed without the proper resources to be treated. Although the clinic has existing partnerships with external mental health facilities, waitlists were long. Furthermore, transportation, insurance, and cost barriers limited patients’ access to mental health treatments. In addition, concerns about the capacity of the current referral network to handle an influx of newly diagnosed patients were expressed by participants.

**CFIR Domain 3: Inner Setting**

Codes within the inner setting domain focused on the clinic’s characteristics and readiness to implement population-level portal-based screening. Participants spoke about compatibility, available resources, access to knowledge and information, networks and communication, and culture.

**Facilitator: Compatibility With Clinic Culture, Workflows, and Systems**

Participants strongly expressed interest in implementing population-level portal-based screening for depression and anxiety through the portal and thought their colleagues would also be receptive. Participants stated that this would help create an open relationship with patients while prioritizing clinical values to provide holistic care to their patients. Participants said that using the portal for depression and anxiety screening would provide crucial information on patients’ mental health status without adding significant stress to clinical workflows.

**Facilitator: Protocol for Addressing Positive Screening Results**

Participants wanted a systematic process to manage patients who screened positive to avoid delays in connecting patients to resources. Specifically, participants expressed a desire for detailed guidance on available resources, referral pathways, and a follow-up plan for symptomatic patients. A few participants were interested in additional training or decision support tools to help them interpret screening results, connect patients to resources, and prescribe first-line medications.

**Barrier: Clinician Variation in the Use of EHRs**

Participants expressed concern that screening results in the EHR might be overlooked because clinicians did not always check their electronic in-baskets reliably because of variations in their proficiency and comfort with the EHR.

**Barrier: Limited Capacity to Address Mental Health Concerns During HIV Visits**

Participants raised concerns about having adequate personnel, time, and expertise to manage depression and anxiety. They reported limited time during appointments to address their patients’ health issues and social needs, and there may not be enough time to address depression and anxiety management.

**CFIR Domain 4: Characteristics of Individuals**

Codes within the individual characteristics domain focused on the participants’ knowledge, beliefs, and self-efficacy.

**Facilitator: Participant Beliefs That Mental Health Screening Is Important and Benefits HIV Care**

Most participants agreed that mental health screening was essential and valuable. They saw population-level portal-based screening as an opportunity to learn more about their patients and to address concerns that might not otherwise arise during appointments. Participants recognized the effects of mental health problems on HIV outcomes and were hopeful that addressing depression and anxiety would improve engagement with care and general health.

**Facilitator: Participant Interest in Evidence-Based Practices and Desire to Learn From Prior Implementations**

Participants indicated their willingness to abide by evidence-based mental health screening and management recommendations. They expressed a desire to learn more about how population-level portal-based depression and anxiety screening had been implemented in the primary care clinic so that the lessons learned could be applied to the HIV clinic.
Barrier: Participant Concerns About Limited Knowledge About Mental Health Treatments

Some participants were hesitant to implement population-level portal-based mental health screening because they thought that they lacked adequate expertise in mental health treatment and navigating mental health resources.

CFIR Domain 5: Process

Codes within the process domain focused on planning the intervention and engaging clinicians and patients.

Facilitator: Team-Based Approach That Leverages the Strengths of All Clinicians

Participants believed that a team approach would be crucial for successfully implementing population-level portal-based screening. Social workers were identified as key team members to provide knowledge on available resources and support connecting patients to care. Participants also indicated that engaging clinicians with strong relationships with patients would help lower patient hesitancy to engage with the portal.

Facilitator: Clear Planning and Communication With Staff

Participants emphasized the importance of having a standardized protocol that included details on which staff member was responsible for each step, especially in response to positive results, and training for all clinic personnel on this protocol before implementation. Participants highlighted the need for clear communication throughout the intervention’s preimplementation, implementation, and sustainability phases. They advised monitoring the intervention logistics and collecting iterative feedback from staff and patients throughout the intervention rollout.

Facilitator: A System That Empowers Patients to Communicate About Their Mental Health

Participants thought that population-level portal-based mental health screening could prompt patients to discuss their mental health with their care team. The participants believed that providing patients with the flexibility to complete screening assessments at their convenience and through their preferred screening method would empower them to inform their care team about their symptoms. Some suggested that screeners should always be readily available in the portal so patients could report their mental health symptoms as they feel them.

Barrier: Impersonal Approach to the Sensitive Topic of Mental Health

Participants expressed concerns that portal-based screening might be impersonal and that unexpected messages about mental health might seem invasive or cause anxiety in some patients. The participants emphasized that clear and patient-centered conversations would need to occur to explain the purpose of mental health screening. Otherwise, the participants feared that patients who did not understand the purpose or context of mental health screening would be unlikely to respond. The participants believed that if patients were informed about how these assessments pertained to their general health, they would be more likely to complete the screeners.

Discussion

Principal Findings

This qualitative study explored facilitators of and barriers to implementing population-level portal-based depression and anxiety screening in an HIV clinic. A total of 10 facilitators and 7 barriers were identified across 5 CFIR domains. Facilitators included the following: advantages of systematic screening outside clinic visits; the expectation that assessment frequency could be tailored to patient needs; evidence from the literature and previous experience in other settings; respect for patient privacy; empowering patients and facilitating communication about mental health; compatibility with clinic culture, workflows, and systems; staff beliefs about the importance of mental health screening and benefits for HIV care; engaging all clinic staff and leveraging their strengths; and clear planning and communication with staff. Barriers included difficulty in ensuring prompt response to suicidal ideation; patient access, experience, and comfort using the portal; limited availability of mental health services; variations in how providers use the EHR and communicate with patients; limited capacity to address mental health concerns during HIV visits; staff knowledge and self-efficacy regarding the management of mental health conditions; and the impersonal approach to a sensitive topic. Several barriers mentioned by participants, such as limited appointment times and limited access to mental health resources after diagnosis, are common challenges cited in similar implementation efforts [2,3,6,7,25,27].

Findings from our analysis have been used to compile a list of proposed implementation strategies to help integrate population-level portal-based depression and anxiety screening into practice within the HIV clinic setting.

Clinician-Focused Implementation Strategies

Strategy 1: Conduct a Local Assessment of Clinicians and Clinic Staff

Clinicians and clinic staff are essential to successfully implementing population-level portal-based depression and anxiety screening in the HIV clinic. To increase the feasibility and sustainability of the intervention, clinicians and clinic staff should be asked how the intervention would fit with their beliefs and values, the clinic culture, and its current clinical workflows. Clinicians’ and clinic staff’s thoughts should be incorporated into the implementation plan to assist in intervention compatibility and uptake.

Strategy 2: Engage Clinicians and Clinic Staff With Various Roles and Expertise to Support Implementation

The success of the intervention depends on clinicians’ engagement through the implementation process to inform the intervention using clinicians and clinic staff’s strengths. Clinicians and clinic staff in various roles may have different perspectives and ideas on implementing the intervention. Therefore, diversifying the staff perspective may provide crucial implementation strategies that might not be known by only interviewing clinicians.
Strategy 3: Highlight Advantages, Relevance, and Evidence for Population-Level Portal-Based Mental Health Screening

Before implementing population-level portal-based mental health screening, the advantages of depression and anxiety screening must be communicated and emphasized to all clinicians. These benefits should highlight how timely mental health discussions between patients and clinicians make efficient use of the limited appointment time. Information on relevant evidence and current clinical screening guidelines should also be provided to garner clinician support. Describing barriers encountered and lessons learned in other practices that have implemented similar interventions could ease concerns about implementation challenges.

Strategy 4: Communicate With Clinicians and Clinic Staff Throughout Implementation and Provide Training on Protocols

Training and involving clinicians throughout the rollout of the intervention will facilitate iterative feedback to troubleshoot any challenges that arise and help aid clinicians and clinic staff uptake. As clinicians and clinic staff tend to have established relationships with their patients, receiving their and their patients’ concerns will aid clinicians, clinic staff, and patient engagement throughout the intervention rollout.

Patient-Focused Implementation Strategies

Strategy 5: Provide Technical Support and Demonstrations for Patients on How to Use the Portal

Providing technical support and conducting training on using the portal might increase intervention uptake among patients. Demonstrations could decrease technology-related barriers and encourage patients to use the portal to complete assessments.

Strategy 6: Use Multiple Screening Methods for Broad Reach

Multiple screening approaches might be needed to reach all patients attending the clinic. For example, options could include completing depression and anxiety screening in the waiting room before an appointment, over the phone, or during an in-person appointment (eg, during intake before the clinician enters the room). Providing additional screening options for patients who are not technologically proficient or have limited access to technology may increase patient uptake of depression and anxiety screening.

Strategy 7: Use Patient-Centered Communication in Portal Messages

Patient-centered messages emphasizing privacy and framing mental health screening as part of routine care can provide a context for portal-based screeners and decrease patient hesitancy to answer questions about the potentially sensitive and stigmatized topic of mental health. Using the patient portal to send patient-centered messages will also allow patients to ask questions about population-level patient-based screening and address concerns.

IT-Focused Implementation Strategies

Strategy 8: Make Screening Frequency Adaptable Based on Patient History and Symptoms

Adaptability of screening frequency and leveraging the staff-patient relationship may improve intervention uptake. Clinicians could adjust the frequency of depression and anxiety screenings based on their relationship with the patient. By allowing staff to adjust the screening frequency, the clinic can check in on patients experiencing uncontrolled depression and anxiety symptoms. Likewise, the staff can lengthen the screening intervals when the patient is in remission for depression and anxiety. This adaptability will signal to patients that the clinic is prioritizing the patient’s health needs.

Strategy 9: Use User-Centered Design Methods to Refine How Results Are Displayed and Communicated in the EHR

When designing how portal-based results will be stored and displayed in the EHR, clinicians and clinic staff need to be engaged to ensure the utility of the screening information. Using a user-centered design with these essential stakeholders could increase the likelihood that portal-based depression and anxiety screening will be used in practice.

Strategy 10: Make Screening Tools Available for Patients to Use on Demand in the Portal

On-demand assessments would support patient autonomy and allow patients to signal when they are experiencing depression and anxiety symptoms. This patient-centered approach could enhance the clinic’s capacity to treat patients when needed. This differs from the traditional annual one-time screening, which aims to identify depression and anxiety in asymptomatic patients. Traditional screening may increase the demand for services and reduce the clinic’s ability to provide timely and appropriate care for symptomatic patients. Moreover, prioritizing on-demand assessments outside of appointment times could facilitate outreach between appointments and reduce the time to treatment.

Clinic-Focused Implementation Strategies

Strategy 11: Create Protocols for Positive Depression and Anxiety Screening Results, Including Those Indicating Imminent Risk

Establishing a standardized protocol for patients who are symptomatic of depression or anxiety may ease concerns about managing patients who are at imminent risk to themselves or others. For example, the protocol can describe who will contact the patient after the clinic has received a positive PHQ or GAD and how to assist patients in crisis. This will ease clinicians’ concerns about screening patients for depression and anxiety via the portal.

Strategy 12: Provide Clinical Decision Support Tools, Training, and Mentorship to Help Clinicians Manage Mental Health Concerns

Providing evidence-based information on treatment or referral strategies through decision support tools, ongoing training, and
clinician mentorship for managing mental health treatment would support clinicians’ confidence and ability to manage symptomatic mental health concerns in their patients. Through shared collaborations with mental health specialists and community mental health services, clinicians will be equipped to manage a potential influx of symptomatic patients via the portal.

**Strategy 13: Implement Integrated Behavioral Health and Increase Mental Health Referral Partnerships**

In concurrence with strategy 12, the clinic will need to invest and establish partnerships with local mental health sites to support the clinic’s capacity to treat newly diagnosed patients. Expanding the clinic’s referral network would create a safety net that the clinic can leverage to refer patients. This would prevent long wait times for treatment after a depression or anxiety diagnosis. By creating a behavioral health referral network, HIV clinicians can provide trusted resources to expand the clinic’s internal infrastructure, facilitate warm handoffs with community partners, and continue to support patient care.

**Limitations**

The study was conducted at a single academic HIV clinic. Therefore, the results of this qualitative analysis may not be generalizable to other HIV clinics with different patient populations, staffing needs, available resources, and portal uptake. The implementation strategies are merely recommendations from a single HIV clinic and may need to be adapted to fit the implementation setting. At the time of the interviews, clinic staff did not have experience with population-level portal-based mental health screening; therefore, their perspectives were based on how they perceived the intervention would be for patients and themselves once implemented. Although our study included perspectives from clinicians in various clinical roles, most interviewees were physicians, limiting available insight. Gaining patient perspectives through patient-focused interviews would provide further insight into facilitators, barriers, and intervention implementation strategies.

**Conclusions**

Our study provides information on clinicians’ views on population-level portal-based mental health screening within the HIV clinic setting. Participating clinicians expressed concerns about the accessibility of prompt mental health resources, patients’ perceptions of mental health screening, variation in clinician use of Epic (Epic Systems, Verona WI), and limited clinician training on mental health management. Nevertheless, clinicians were interested in establishing population-level portal-based screening at the HIV clinic and were amenable to creating protocols for addressing positive mental health screening, to participating in training about available mental health resources and best practices, and to feeling it was compatible with the clinic. Others may build upon this work by exploring and identifying additional facilitators, barriers, and implementation strategies that were not found in our analysis.

**Acknowledgments**

This publication was made possible with support from the Third Coast Center for AIDS Research, a National Institutes of Health–funded center (P30 AI117943).

**Authors’ Contributions**

EMS, JPR, JS, MF, SJH, DM, and NL contributed to study concept and design. Data acquisition was performed by MF. Data analysis and interpretation were performed by DZ, EMS, and NL. DZ wrote the initial manuscript draft. Critical revision of the manuscript for intellectual content was performed by all the authors.

**Conflicts of Interest**

JPR has received fees for legal consulting from Gilead Sciences.

**References**


https://formative.jmir.org/2024/1/e48935


Abbreviations

CFIR: Consolidated Framework for Implementation Research
EHR: electronic health record
GAD: Generalized Anxiety Disorder
PHQ: Patient Health Questionnaire
REDCap: Research Electronic Data Capture

©Daniela Zimmer, Erin M Staab, Jessica P Ridgway, Jessica Schmitt, Melissa Franco, Scott J Hunter, Darnell Motley, Neda Laiteerapong. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Dentists’ Information Needs and Opinions on Accessing Patient Information via Health Information Exchange: Survey Study

Shuning Li1, PhD; Grace Gomez Felix Gomez1,2, BDS, PhD; Huiping Xu3, PhD; Anushri Singh Rajapuri1, BDS, MS; Brian E Dixon2,4, PhD; Thankam Thyvalikakath1,2, DMD, MDS, PhD

1Department of Dental Public Health and Dental Informatics, Indiana University School of Dentistry, Indianapolis, IN, United States
2Regenstrief Institute, Inc, Center for Biomedical Informatics, Indianapolis, IN, United States
3Department of Biostatistics and Health Data Sciences, Indiana University School of Medicine, Indianapolis, IN, United States
4Department of Epidemiology, Indiana University Richard M Fairbanks School of Public Health, Indianapolis, IN, United States

Corresponding Author:
Thankam Thyvalikakath, DMD, MDS, PhD
Department of Dental Public Health and Dental Informatics
Indiana University School of Dentistry
415 Lansing St
Indianapolis, IN, 46201
United States
Phone: 1 3172745460
Email: tpt@iu.edu

Abstract

Background: The integration of medical and dental records is gaining significance over the past 2 decades. However, few studies have evaluated the opinions of practicing dentists on patient medical histories. Questions remain on dentists’ information needs; their perception of the reliability of patient-reported medical history; satisfaction with the available information and the methods to gather this information; and their attitudes to other options, such as a health information exchange (HIE) network, to collect patient medical history.

Objective: This study aims to determine Indiana dentists’ information needs regarding patients’ medical information and their opinions about accessing it via an HIE.

Methods: We administered a web-based survey to Indiana Dental Association members to assess their current medical information-retrieval approaches, the information critical for dental care, and their willingness to access or share information via an HIE. We used descriptive statistics to summarize survey results and multivariable regression to examine the associations between survey respondents’ characteristics and responses.

Results: Of the 161 respondents (161/2148, 7.5% response rate), 99.5% (n=160) respondents considered patients’ medical histories essential to confirm no contraindications, including allergies or the need for antibiotic prophylaxis during dental care and other adverse drug events. The critical information required were medical conditions or diagnosis, current medications, and allergies, which were gathered from patient reports. Furthermore, 88.2% (n=142) of respondents considered patient-reported histories reliable; however, they experienced challenges obtaining information from patients and physicians. Additionally, 70.2% (n=113) of respondents, especially those who currently access an HIE or electronic health record, were willing to use an HIE to access or share their patient’s information, and 91.3% (n=147) shared varying interests in such a service. However, usability, data accuracy, data safety, and cost are the driving factors in adopting an HIE.

Conclusions: Patients’ medical histories are essential for dentists to optimize dental care, especially for those with chronic conditions. In addition, most dentists are interested in using an HIE to access patient medical histories. The findings from this study can provide an alternative option for improving communications between dental and medical professionals and help the health information technology system or tool developers identify critical requirements for more user-friendly designs.

(JMIR Form Res 2024;8:e51200) doi:10.2196/51200

KEYWORDS
dentistry; medical history; integrated medical and dental records; health information exchange; medical record; dental record; dental; medical information; dental care; adverse drug effect; medication; allergies; cost; data safety; data accuracy
Introduction

More than 20 years ago, the first US Surgeon General’s Report on Oral Health in America established oral health as an essential component of overall health and well-being [1]. In 2021, the National Institutes of Health (NIH) reemphasized the importance of establishing integrated medical and dental care in their updated report on Oral Health in America [2]. In addition, the NIH identified integrating medical and dental records as critical to enhancing medical and dental care [2]. The integration of medical and dental records is gaining significance for several reasons. First, increased evidence during the last 3 decades indicates strong associations and shared risk factors between oral and systemic diseases such as diabetes and heart diseases [2,3]. Second, the siloed systems of dental and medical data create challenges in information sharing [2,4,5], often resulting in incomplete or inaccurate patient medical information, which may cause significant patient care and safety issues in dental care [6-9]. Third, recent studies have revealed discrepancies in medical conditions and medications in the electronic dental record (EDR) versus electronic health record (EHR) [3,6,10-14] and demonstrated substantial delays when dentists are required to request additional medical information from physicians [15]. Other reasons for the increased significance of integrated medical and dental records include the rapid development of information technologies, which provides a solid base for integration, and the impact of the COVID-19 pandemic, which proves the urgency and importance of integrating medical and dental records [16-18].

The use of an integrated EDR-EHR system has been growing in large health care organizations (HCOs) such as the Veteran Affairs health care systems, Department of Defense, health maintenance organizations, and federally qualified health centers, where medical and dental practices are colocated and share patient care and records [19-21]. Numerous studies have reported physicians using integrated EDR-EHR systems to refer patients to dentists and vice versa for preventive and comprehensive care [19,21-23]. However, for most dentists who work in small independent practices, patients continue to be the primary source of their medical history and dentists’ communications with medical providers are limited [24]. Nevertheless, the solo and small-group dental practices, which constitute 50% of the dental workforce [25], cannot adopt such integrated systems without being credentialed to a major HCO. It is also not practical for dental practices to have separate interfaces to different EHR systems, which may interfere with their clinical workflows and business processes, such as billings and regulatory policies [3].

With the support of several federal policies, such as the Health Information Technology for Economic and Clinical Health Act and the 21st Century Cures Act [26,27], and the financial incentives established by the Centers for Medicare and Medicaid Services [28,29], community and regional health information exchanges (HIEs) have expanded significantly since 2009. HIE systems provide another option for integrating medical and dental records [24]. In an earlier study by the team, we modeled 3 methods for dentists to access their patient’s medical histories: the patient-reported medical history followed by the optional medical consults method, the integrated EDR-EHR, and the HIE approaches [24]. Our models showed that the HIE approach could provide benefits for reducing unnecessary medical consultations, avoiding the delay of care, improving information quality, and cutting additional technical and financial overheads for small independent practices. In addition, a report published in 2021 indicated a decrease in dentists working solo [25]. Nevertheless, an HIE-based integrated solution can help small and large group practices improve data completeness and compliance by obtaining data from multiple HCOs and taking advantage of expert services provided by an HIE. However, efforts to connect dentists with an HIE are minimal compared to the extensive studies on integrated EDR-EHR [19-22].

Despite the widespread interest in integrating dental and medical care, few studies have evaluated the opinions of practicing dentists on patient medical histories [26,30]. For example, a recent study published by the American Dental Association Clinical Evaluators Panel reported that most dentists gathered their patients’ medical history and medication list via patients’ self-report and recorded vital signs during dental visits [30]. However, questions remain on dentists’ information needs; their perception of the reliability of patient-reported medical history; satisfaction with the available information and the methods to gather this information; and their attitudes to other options, such as an HIE, to collect patient medical history.

Given this knowledge gap, we surveyed dentists in 1 US state about their information needs and practices concerning retrieving patient medical history information. Our objectives were to determine their information needs regarding patients’ medical information and their opinions on accessing it via an HIE.

Methods

Recruitment

We administered a web-based survey to the Indiana Dental Association (IDA) members from March 19, 2021, to April 30, 2021. All participants are general dentists or specialists either currently or previously practicing in the State of Indiana. We only included dentists in this survey because they are responsible for diagnosing and planning treatments, which also involve ruling out contraindications. The survey was administrated through the Indiana University–approved Qualtrics Experience Management platform XM. We sent emails to 2148 IDA members over 6 weeks, including 1 initial invitation, 5 reminders, and 1 final thank you note.

Ethical Considerations

Participation in the survey was voluntary, responses were anonymous, and participants could only respond once (configurations blocked multiple responses in the web-based survey tool). Participants gave informed consent by accessing the link provided in the study invitation email. The patients were not compensated. This study received exemption approval from the Indiana University Institutional Review Board (Protocol #2012972646).
Survey Construction and Validation
The survey included 27 questions covering 3 topics: 12 on demographics, 11 on information needs and gathering, and 4 on exchanging patient medical information (Multimedia Appendix 1). Demographic information included sex, years in practice, primary practice information (type of practice, general practitioner or specialist, typical procedures, patient age groups), and EDR use. The information needs and data gathering section included questions related to dentists’ information needs, existing methods for collecting information, and challenges in these approaches. Finally, the exchange of patient medical information section included dentists’ opinions on using an HIE-based information platform to receive and share patients’ medical information. The survey had 23 multiple-choice questions, two 0-10 Likert-scale questions, and 2 open-ended questions. We administered the survey after assessing the face validity and content validity of the questionnaire. The face validity was assessed with research team members who were not involved in the development of the survey questions, and the content validity was assessed with 3 dentists—2 from the Indiana University School of Dentistry and 1 from private practice. These tests ensured that the survey was appropriate, understandable, and could be completed within a reasonable time.

Statistical Analysis
Data analysis included only completed responses. Partially answered surveys were eliminated from the final analysis. Data were summarized using frequencies and percentages for categorical variables and mean and SDs for continuous variables.

Associations between characteristics of dentists (years in practice, dental professions, and current access to an HIE or a hospital or medical practice-based EHR [hereby referred to as HIE-EHR]) and their opinions on the importance and reliability of patients’ medical histories and perceptions of accessing patients’ medical histories via an HIE were examined using multivariable regression. The ordinal logistic regression model was used due to the ordinal nature of these variables. All statistical analyses were performed using SAS (version 9.4, The SAS Institute). P<.05 were considered statistically significant.

Results
At the end of 6 weeks, 219 (10.2%) out of 2148 IDA members accessed the survey, and 188 (8.8%) members responded to at least 1 question, of which 161 (7.5%) members reached the end of the survey.

Demographics
A total of 64.6% (102/158) of the respondents were male (Table 1). Their average years in practice was 25.72 (SD 13.52) years. A total of 74.5% (120/161) of the respondents were general practitioners, and the rest were dental specialists (Table 1). A total of 8 dental specialties were reported: oral and maxillofacial surgery (10/41, 24.2%), periodontics (8/41, 19%), orthodontics (7/41, 16.7%), pediatric dentistry (7/41, 16.7%), endodontics (4/41, 11.9%), operative dentistry (2/41, 4.8%), prosthodontics (2/41, 4.8%), and oral and maxillofacial pathology (1/41, 2.4%). The total percentage is more than 100% since some respondents reported more than 1 specialty.
Table 1. Characteristics of survey respondents.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Survey respondents</strong></td>
<td></td>
</tr>
<tr>
<td>Sex (n=158), n (%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>56 (35.4)</td>
</tr>
<tr>
<td>Male</td>
<td>102 (64.6)</td>
</tr>
<tr>
<td>Dental profession (n=161), n (%)</td>
<td></td>
</tr>
<tr>
<td>General practitioner</td>
<td>120 (74.5)</td>
</tr>
<tr>
<td>Dental specialist</td>
<td>41 (25.5)</td>
</tr>
<tr>
<td>Years in practice (n=161), mean (SD)</td>
<td>25.72 (13.52)</td>
</tr>
<tr>
<td><strong>Survey respondents’ primary dental practices</strong></td>
<td></td>
</tr>
<tr>
<td>Type (n=160), n (%)</td>
<td></td>
</tr>
<tr>
<td>Private dental practice owner</td>
<td>100 (62.5)</td>
</tr>
<tr>
<td>Associate dentist of a private practice</td>
<td>26 (16.3)</td>
</tr>
<tr>
<td>Major dental care organizations such as dental schools and health maintenance organizations</td>
<td>19 (11.9)</td>
</tr>
<tr>
<td>Public health practice, community health center, or publicity-funded clinic (but not a federal facility)</td>
<td>2 (1.3)</td>
</tr>
<tr>
<td>Federal government facility (Veterans Affairs, Department of Defense, and Public Health Service)</td>
<td>2 (1.3)</td>
</tr>
<tr>
<td>Other</td>
<td>11 (6.9)</td>
</tr>
<tr>
<td>Number of dentists (including the respondent; n=160), n (%)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>75 (46.9)</td>
</tr>
<tr>
<td>2-5</td>
<td>66 (41.3)</td>
</tr>
<tr>
<td>6-10</td>
<td>3 (1.9)</td>
</tr>
<tr>
<td>&gt;10</td>
<td>16 (10)</td>
</tr>
<tr>
<td>Number of hygienists (n=161), n (%)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>31 (19.3)</td>
</tr>
<tr>
<td>1-5</td>
<td>110 (68.3)</td>
</tr>
<tr>
<td>6-10</td>
<td>15 (9.3)</td>
</tr>
<tr>
<td>&gt;10</td>
<td>5 (3.1)</td>
</tr>
<tr>
<td>Use EDR&lt;sup&gt;a&lt;/sup&gt; to manage clinical data (n=161), n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>128 (79.5)</td>
</tr>
<tr>
<td>No</td>
<td>33 (20.5)</td>
</tr>
<tr>
<td>EDR system brands (n=126), n (%)</td>
<td></td>
</tr>
<tr>
<td>Dentrix</td>
<td>45 (35.7)</td>
</tr>
<tr>
<td>EagleSoft</td>
<td>19 (15.1)</td>
</tr>
<tr>
<td>axiUm</td>
<td>14 (11.1)</td>
</tr>
<tr>
<td>OpenDental</td>
<td>10 (7.9)</td>
</tr>
<tr>
<td>SoftDent</td>
<td>10 (7.9)</td>
</tr>
<tr>
<td>Practice Works</td>
<td>6 (4.8)</td>
</tr>
<tr>
<td>Easy Dental</td>
<td>3 (2.4)</td>
</tr>
<tr>
<td>Other&lt;sup&gt;b&lt;/sup&gt;</td>
<td>19 (15.1)</td>
</tr>
<tr>
<td>Whether or not have access to a state-based health information exchange, exchange capability between dental software and electronic medical record system, or integrated dental-medical record system? (n=161), n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>24 (14.9)</td>
</tr>
<tr>
<td>No</td>
<td>137 (85.1)</td>
</tr>
</tbody>
</table>
Patient age distribution (%; n=157), mean (SD)

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 years and younger</td>
<td>19.6 (23.3)</td>
</tr>
<tr>
<td>19-44 years</td>
<td>26.1 (13.8)</td>
</tr>
<tr>
<td>45-64 years</td>
<td>31.8 (15)</td>
</tr>
<tr>
<td>65 years and older</td>
<td>22.5 (13.1)</td>
</tr>
</tbody>
</table>

*EDR: electronic dental record.

Other EDR brands included Ascend (by Dentrix), Cloud9, Curve Dental, DentiCon, DRM plus, DSN PerioExec, EPMS, MacPractice, Mconsent, Florida Probe, Mogo, NextGen, OMSvision, Ortho2 Edge, Practice Fusion, and Practice Web.

In all, 78.8% (126/160) of respondents reported working in private practices as owners or associate dentists. Approximately half (75/160, 46.9%) of the respondents reported having 1 dentist in their primary dental practices, while 10% (16/160) of respondents reported their primary practice having more than 10 dentists. Most respondents’ (110/160, 68.3%) primary practices had 2-5 hygienists, while 19% (31/160) of respondents’ practices had no hygienists. The 3 most frequently performed procedures were diagnostic and preventive such as an examination, X-rays, scaling, prophylaxis, sealants, fluoride, etc (136/161, 84.4%); restorations or fillings (125/161, 78.1%); and tooth-supported or implant-supported crowns (105/161, 65.3%). The respondents served a diverse age group of the patient population, with an average of 19.6% (31/157) of patients 18 years or younger and 22.5% (35/157) of patients 65 years or older.

About 4 in 5 (128/161, 79.5%) respondents reported using an EDR for not only billing or scheduling but also for clinical or patient data management. The top 3 brands of EDR were Dentrix (145/161, 35.7%), EagleSoft (19/161, 15.1%), and aXiUm (14/161, 11.1%; Table 1).

### Dentists’ Opinions on the Importance of Medical Histories and Reliability of Patient-Reported Medical Histories

Almost all respondents (160/161, 99.5%) considered patients’ medical histories highly or moderately important during dental care. They reviewed medical histories to (1) verify no contraindications exist to undergo a dental procedure (37/161, 23.2%), (2) confirm no need for antibiotic prophylaxis before the dental procedure (36/161, 22.7%), (3) rule out any allergies or adverse drug reactions (35/161, 22%), (4) assist with determining the prognosis of an oral disease or treatment outcomes (35/161, 21.7%), (5) detect normal and abnormal laboratory results (14/161, 8.4%), and (6) for other purposes (3/161, 1.9%).

Regarding the reliability of patient-reported medical histories, 8% (n=13) of respondents considered them highly reliable, 79.5% (n=128) moderately reliable, and 12.4% (n=20) unreliable.
Table 2. Respondents’ opinions on the importance and reliability of patients’ medical histories and their perceptions of accessing patient history via an HIE\(^a\) (n=161).

<table>
<thead>
<tr>
<th>Opinions</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Opinions to patients’ medical histories(^b)</strong></td>
<td></td>
</tr>
<tr>
<td>How important is obtaining patient’s up-to-date medical history for you?</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>7</td>
<td>3 (1.9)</td>
</tr>
<tr>
<td>8</td>
<td>13 (8.1)</td>
</tr>
<tr>
<td>9</td>
<td>18 (11.2)</td>
</tr>
<tr>
<td>10: Extremely important</td>
<td>126 (78.3)</td>
</tr>
<tr>
<td>How reliable is the patient-reported medical history?</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2 (1.2)</td>
</tr>
<tr>
<td>3</td>
<td>3 (1.9)</td>
</tr>
<tr>
<td>4</td>
<td>4 (2.5)</td>
</tr>
<tr>
<td>5</td>
<td>11 (6.8)</td>
</tr>
<tr>
<td>6</td>
<td>29 (18.0)</td>
</tr>
<tr>
<td>7</td>
<td>57 (35.4)</td>
</tr>
<tr>
<td>8</td>
<td>42 (26.1)</td>
</tr>
<tr>
<td>9</td>
<td>11 (6.8)</td>
</tr>
<tr>
<td>10: Extremely reliable</td>
<td>2 (1.2)</td>
</tr>
<tr>
<td><strong>Perceptions of accessing patient history via an HIE</strong></td>
<td></td>
</tr>
<tr>
<td>Do you think access to such a system would be useful?</td>
<td></td>
</tr>
<tr>
<td>1: No</td>
<td>8 (5.0)</td>
</tr>
<tr>
<td>2: Maybe</td>
<td>40 (24.8)</td>
</tr>
<tr>
<td>3: Yes</td>
<td>113 (70.2)</td>
</tr>
<tr>
<td>Would you consider using it to access your patient’s medical information?</td>
<td></td>
</tr>
<tr>
<td>1: No</td>
<td>9 (5.6)</td>
</tr>
<tr>
<td>2: Maybe</td>
<td>40 (24.8)</td>
</tr>
<tr>
<td>3: Yes</td>
<td>112 (69.6)</td>
</tr>
<tr>
<td>Would you allow other health care providers to access clinical information about your own patients?</td>
<td></td>
</tr>
<tr>
<td>1: No</td>
<td>15 (9.3)</td>
</tr>
<tr>
<td>2: Maybe</td>
<td>47 (29.2)</td>
</tr>
<tr>
<td>3: Yes</td>
<td>99 (61.5)</td>
</tr>
<tr>
<td>What is your interest to participate in a service to access such as a system?</td>
<td></td>
</tr>
<tr>
<td>1: Not interested at all</td>
<td>15 (9.3)</td>
</tr>
<tr>
<td>2: Slightly interested</td>
<td>19 (11.8)</td>
</tr>
<tr>
<td>3: Moderately interested</td>
<td>58 (36.0)</td>
</tr>
<tr>
<td>4: Very interested</td>
<td>39 (24.2)</td>
</tr>
<tr>
<td>5: Extremely interested</td>
<td>30 (18.6)</td>
</tr>
</tbody>
</table>

\(^a\)HIE: health information exchange.
\(^b\)10-level Likert scale was used with ranges as follows: 1-5=not important or reliable, 6-8=moderately important or reliable, and 9-10=highly important or reliable.
Dentists’ Information Needs Regarding Their Patient’s Medical History

The 3 most needed information categories for new and existing patients were medical conditions or diagnosis, current medications, and allergies. Other categories included hospitalizations in the last 2 years, substance abuse, procedures in the previous 5 years, laboratory results from the last 6 months, and immunization records. The respondents evaluated the information needs of new and existing patients separately, and there were no significant differences in the results (Table 3).

**Table 3.** Dentists’ most needed patient medical information during dental care.

<table>
<thead>
<tr>
<th>Patient medical information</th>
<th>Existing patients (n=161), n (%)</th>
<th>New patients (n=161), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical condition or diagnoses</td>
<td>150 (93.2)</td>
<td>153 (95)</td>
</tr>
<tr>
<td>Current medications</td>
<td>141 (87.6)</td>
<td>148 (91.9)</td>
</tr>
<tr>
<td>Allergies</td>
<td>138 (85.7)</td>
<td>143 (88.8)</td>
</tr>
<tr>
<td>Substance abuse</td>
<td>76 (47.2)</td>
<td>81 (50.3)</td>
</tr>
<tr>
<td>Hospitalization in the last 2 years</td>
<td>70 (43.5)</td>
<td>68 (42.2)</td>
</tr>
<tr>
<td>Procedures in the last 5 years</td>
<td>44 (27.3)</td>
<td>49 (30.4)</td>
</tr>
<tr>
<td>Laboratory results from the last 6 months</td>
<td>17 (10.6)</td>
<td>17 (10.6)</td>
</tr>
<tr>
<td>Immunization records</td>
<td>4 (2.5)</td>
<td>3 (1.9)</td>
</tr>
<tr>
<td>Others</td>
<td>3 (1.9)</td>
<td>5 (3.1)</td>
</tr>
</tbody>
</table>

**Dentists’ Access to Their Patient’s Medical History**

We also asked the dentists how they collected patient-reported medical histories and obtained additional information if needed. Paper-based health history forms constituted the most used method (127/161, 78.9%), followed by web-based health history forms (62/161, 38.5%) and electronic devices such as tablets (35/161, 21.7%). The total percentage is more than 100% since some respondents reported using more than 1 method. The top 3 challenges in collecting patient-reported medical history were as follows: (1) patients do not remember or recall medication names and dosage (156/161, 96.9%); (2) patients do not recall previous procedures and medical conditions (129/161, 80.1%); and (3) patients’ reluctance to share their medical history (84/161, 52.2%). When the respondents needed additional information, most (158/161, 98%) contacted physicians’ offices or health care providers directly via phone, fax, or email. Other communication methods included paper-based medical consult forms through the patient (46/161, 28.6%), patient’s pharmacy (39/161, 24.2%), state-based HIE (19/161, 11.8%), exchange capability between dental software and electronic medical record system (5/161, 3.1%), integrated dental-medical record system (4/161, 2.5%), and other (9/161, 5.6%). However, they experienced challenges such as the need for multiple attempts (97/161, 60.2%), not receiving information on time (80/161, 49.7%), physician offices being nonresponsive (66/161, 41.0%), need to contact numerous providers or specialists (55/161, 34.2%), need for patient intervention (44/161, 27.3%), and not receiving requested information (35/161, 21.7%).

**Dentists’ Perceptions of Accessing Patient History via an HIE**

A total of 69.6% (113/161) of respondents considered access to a regional HIE useful (Table 2). If such a system were available, 69.9% (n=113) of the respondents would consider using it to access their patient’s medical information, and 61.5% (n=99) would be willing to allow other healthcare providers to access their patients’ clinical information (Table 2).

Furthermore, 91.3% (n=147) of the respondents expressed various interests in participating in a service to access an HIE (Table 2). However, they expressed concerns over the design and implementation of such a system, including data accuracy, data security and HIPAA (Health Insurance Portability and Accountability Act) compliance, cost of implementation (both time and money), and system usability.

The association between respondent characteristics (including dental profession, number of years in practice, and current access to an EHR or HIE) and their opinions on the importance and reliability of patients’ medical histories and perceptions of accessing patients’ medical histories via HIE based on multivariable ordinal logistic regression is displayed in Table 4. Dental profession type (general practitioner vs dental specialist) does not significantly affect one’s opinions toward the importance (P=.98) and reliability (P=.31) of patients’ medical history. However, respondents with more than 40 years in practice were less likely to consider obtaining up-to-date patient information important compared to those with less than 40 years in practice (odds ratio [OR] 0.351, 95% CI 0.139-0.889; P=.047). In addition, respondents with access to an HIE-EHR were more likely to consider obtaining up-to-date patient information important compared to those who do not have access to an HIE-EHR (OR 2.267, 95% CI 1.011-5.084; P=.047). Regarding the respondents’ perceptions of using an HIE to access patients’ medical histories, we found that dental specialists were more interested than general practitioners in participating in service to access patient information via an HIE (OR 2.267, 95% CI 1.174-4.378; P=.02). Compared to respondents without current access to an HIE-EHR, those with access to an HIE-EHR were more likely to think it worthwhile to access such a system (OR 6.306, 95% CI 2.671-14.886; P<.001), more likely to consider using such a system to access their patient’s information (OR 5.538, 95% CI 2.379-12.892; P<.001), more likely to allow other providers to access their patient’s data (OR 2.943, 95% CI 1.342-6.456;
In this study, we determined community practice dentists’ and medical care in large health care systems [19-21]. Through existing studies only highlight case studies of integrating dental has gained tremendous attention in recent years [2]. However, care coordination between dental and medical providers and medical record data is critical to promote communication and dental providers’ use of such services. Integration of dental and accuracy, data safety, and implementation costs would drive Nevertheless, the participants commented that usability, data accuracy, data safety, and implementation costs would drive dental providers’ use of such services. Integration of dental and medical record data is critical to promote communication and care coordination between dental and medical providers and has gained tremendous attention in recent years [2]. However, existing studies only highlight case studies of integrating dental and medical care in large health care systems [19-21]. Through this study, we determined community practice dentists’ information needs and attitudes toward accessing patient medical information via an HIE. These study results contribute to dental professionals’ high-priority information needs and HIE functionalities for successfully using the expanding HIE network in the United States and other countries. In the sections below, we discuss the relevant findings in detail.

Dentists with <40 years of experience or having access to an HIE-EHR system felt patients’ medical histories were more critical than those with >40 years of experience, even though almost 90% (145/161) of the dentists considered patients’ medical histories essential (Table 2). This difference could be because, until 2 decades ago, only limited information technology existed for dentists to access their patient’s medical information except for patient-reported medical history and medical consults. This limited access to EHR data may explain why dentists with more than 40 years in practice were more likely to think patient-reported information as reliable (Table 4). Additionally, dentist respondents who already have access to an HIE-EHR system may benefit more from their patients’ medical histories since they have easier access to the information and may have better quality of information.

Our survey found that the most needed information categories were medical conditions or diagnosis, current medications, and allergies (Table 3), which was consistent with a previous survey [26]. Together these findings showed that some categories of patient medical information were more helpful to dentists during dental care. These findings can also be used to optimize the user interface design in either an EDR-EHR system or an HIE to avoid information overload. However, our team’s earlier studies discovered that dentists’ most requested information categories were laboratory values and written diagnostic reports, followed by recommendations or medical

### Discussion

#### Principal Findings

We surveyed Indiana dentists to determine their information needs regarding patients’ medical histories and their opinions of accessing patient-specific medical information via a community or regional HIE. The survey respondents’ demographics distribution closely matched the dentists’ demographics in the 2020 Indiana oral health workforce data report [31]. In addition, the response rate of 7.5% (161/2148) is comparable to previous surveys of health care professionals, especially web-based surveys [32-34]. The results demonstrated dentists’ high priority in obtaining their patients’ medical diagnoses or conditions, medication histories, and allergies to provide optimum dental care. The survey respondents also reported challenges in getting medical information from patients and medical providers, although they considered patient-reported medical histories moderately or highly reliable. It is also significant that 70% (112/160) of surveyed dentists who work primarily in community practices (Table 1) expressed willingness to use and participate in a service to access and share their patients’ medical histories via an HIE.

Nevertheless, the participants commented that usability, data accuracy, data safety, and implementation costs would drive dental providers’ use of such services. Integration of dental and medical record data is critical to promote communication and care coordination between dental and medical providers and has gained tremendous attention in recent years [2]. However, existing studies only highlight case studies of integrating dental and medical care in large health care systems [19-21]. Through this study, we determined community practice dentists’ information needs and attitudes toward accessing patient medical information via an HIE. These study results contribute to dental professionals’ high-priority information needs and HIE functionalities for successfully using the expanding HIE network in the United States and other countries. In the sections below, we discuss the relevant findings in detail.

#### Table 4. Impact of respondents’ demographics on their opinions on patient medical history and perceptions on accessing patient medical information via an HIE.

<table>
<thead>
<tr>
<th>Patient medical information</th>
<th>Dental specialist vs general practitioner</th>
<th>&gt;40 vs ≤40 years in practice</th>
<th>Have access to HIE-EHR&lt;sup&gt;b&lt;/sup&gt; vs no access</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95% CI)</td>
<td>P value</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td>How important is obtaining a patient’s up-to-date medical history for you?</td>
<td>0.988 (0.401-2.437)</td>
<td>.98</td>
<td>0.351 (0.139-0.889)</td>
</tr>
<tr>
<td>How reliable is the patient-reported medical history?</td>
<td>0.713 (0.374-1.360)</td>
<td>.31</td>
<td>2.267 (1.011-5.084)</td>
</tr>
<tr>
<td>Do you think access to such a system would be useful?</td>
<td>1.567 (0.674-3.643)</td>
<td>.30</td>
<td>2.435 (0.821-7.217)</td>
</tr>
<tr>
<td>Would you consider using it to access your patient’s medical information?</td>
<td>2.187 (0.908-5.264)</td>
<td>.08</td>
<td>1.577 (0.577-4.309)</td>
</tr>
<tr>
<td>Would you allow other providers to access clinical information about your own patients?</td>
<td>1.311 (0.623-2.759)</td>
<td>.48</td>
<td>1.517 (0.602-3.825)</td>
</tr>
<tr>
<td>What is your interest in participating in a service to access such a system?</td>
<td>2.267 (1.174-4.378)</td>
<td>.02&lt;sup&gt;d&lt;/sup&gt;</td>
<td>1.609 (0.722-3.585)</td>
</tr>
</tbody>
</table>

<sup>a</sup>HIE: health information exchange.
<sup>b</sup>EHR: electronic health record.
<sup>c</sup>OR: odds ratio.
<sup>d</sup>P<.05 were considered statistically significant.
clearances [15]. The inconsistency of these results indicates that dentists’ information needs can evolve based on access to relevant information. As they gain access to EHR information, they can ask more specific and informed questions when consulting their medical colleagues, leading to increased responses from medical colleagues. This improved information access may enhance dentists’ patient management and treatment planning. The results also indicated dentists’ information needs for new and existing patients were almost identical (Table 3). Future studies should continue investigating dentists’ information needs as they gain direct access to patients’ up-to-date medical information via an EDR-EHR system or an HIE.

The survey respondents, especially those with access to an HIE-EHR, showed clear interest in using the HIE to optimize the information collection process (Tables 2 and 3). For instance, 11.8% (19/161) of the respondents reported access to a state-based HIE, which was higher than expected. This higher access rate could be attributed to dental providers’ access to state-wide information systems, such as Indiana’s Prescription Drug Monitoring Program, and may have mistaken it for an HIE. Nonetheless, several state-wide HIEs are promoting dentists’ use of HIEs to improve access to patient information [35-37]. However, the overall use during dental care remains low. For instance, a study of New York dentists’ use of the Rochester regional HIE demonstrated a 0.17% rate of use of the HIE during dental encounters [35]. This low use is not surprising given that the use of community HIEs, even by non-dental providers, is still growing, with 1% to 5% use in all patient visits. In the New York dentists’ study, they accessed the HIE primarily for patients with chronic conditions, gingival and periodontal diagnosed diseases, and during the first dental visit [35]. The most frequently visited sections were the laboratory and radiology sections within the HIE, which is consistent with our earlier study results of dentists’ medical consult requests [15].

Although the emergence of community and vendor-supported HIEs has improved medical providers’ timely access to patient information [38-41], inefficient and cumbersome processes and poor user experiences are significant barriers to HIE use [42,43]. Previous studies in medical settings reported that some HIEs require users to have multiple logins; interrupt their workflow; and display overloaded and poorly arranged information [42,43]. Our study respondents expressed similar concerns about the usability of HIEs, such as difficulty accessing data, information overload, and nonintuitive interface designs that could prevent dentists’ use of HIEs. Therefore, future HIE tools’ design and development should focus on the accuracy and integration of the data (content) and the information display and navigation (presentation). Few respondents including those willing to use an HIE expressed concerns about accessing patients’ medical histories via an HIE due to data safety and HIPAA compliance concerns. This issue needs to be addressed both at the technical level with more new tools and methods to ensure safe data sharing and exchanging and at the regulatory level with new protocols and rules to support the use of HIEs. Furthermore, most respondents agreed that patients should be able to control the use of their health care information, and their consent must be received before any information exchange and sharing occur.

Limitations
This study only invited Indiana dentists who are IDA members. A more geographically diversified pool of participants may help improve the results’ validity and generalizability. In future studies, we also want to include other dental professionals, such as dental hygienists and dental assistants. We are aware of the relatively low response rate to the survey, which is not rare in surveys of health care professionals, especially web-based surveys [32-34]. This was an exploratory study and our initial step to determine dentists’ information needs and to help improve their information access. Based on the results of this survey, we will conduct key informant interviews and focus group studies to include a broader group of participants. Another limitation was that dentists may not be familiar with some of the terminologies used in the survey such as state-based HIEs. Terminology definitions and examples should be included in future survey designs.

Conclusions
Patients’ medical histories are essential for dentists to provide high-quality dental care. In addition, information such as medical conditions or diagnosis, current medications, and allergies are more relevant to dentists’ clinical decision-making. Paper-based health history forms and medical consults are still the most widely used methods to gather information. However, electronic forms and integrated systems are gaining attention to have direct access to information. Most dentists are interested in using an HIE to access patient medical histories. The findings from this study can provide an alternative option for improving communications between dental and medical professionals and help the health information technology system or tool developers identify critical requirements for more user-friendly designs.

Acknowledgments
This study was supported by the National Institute of Dental and Craniofacial Research of the National Institutes of Health under award 5R01DE031259-02. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Institute of Dental and Craniofacial Research.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request. Requests may need approval from Indiana University.
Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey on how essential patient medical history is for dental care.

[PDF File (Adobe PDF File), 152 KB - formative_v8i1e51200_app1.pdf]

References


Abbreviations

EDR: electronic dental record
EHR: electronic health record
HCO: health care organization
HIE: health information exchange
HIPAA: Health Insurance Portability and Accountability Act
IDA: Indiana Dental Association
NIH: National Institutes of Health
OR: odds ratio

©Shuning Li, Grace Gomez Felix Gomez, Huiping Xu, Anushri Singh Rajapuri, Brian E Dixon, Thankam Thyvalikakath. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
MIR FORMATIVE RESEARCH

Original Paper

Promoting Self-Efficacy of Individuals With Autism in Practicing Social Skills in the Workplace Using Virtual Reality and Physiological Sensors: Mixed Methods Study

Sung-In Kim¹, MD; So-youn Jang², MS; Taewan Kim³, PhD; Bogoan Kim⁴, BA; Dayoung Jeong⁵, BS; Taehyung Noh⁵, BS; Mingon Jeong⁵, BS; Kaely Hall⁶, MS; Meelim Kim⁷,8,9,10, PhD; Hee Jeong Yoo¹¹,12, MD, PhD; Kyungsik Han¹⁴,5, PhD; Hwajung Hong¹⁵*, PhD; Jennifer G Kim²*, PhD

¹Department of Psychiatry, Seoul National University Bundang Hospital, Bundang, Republic of Korea
²Georgia Institute of Technology, Atlanta, GA, United States
³Department of Industrial Design, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea
⁴Department of Data Science, Hanyang University, Seoul, Republic of Korea
⁵Department of Artificial Intelligence, Hanyang University, Seoul, Republic of Korea
⁶School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, United States
⁷Department of Preventive Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea
⁸Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, San Diego, CA, United States
⁹The Design Lab, University of California San Diego, San Diego, CA, United States
¹⁰Center for Wireless & Population Health Systems Calit2's Qualcomm Institute, University of California San Diego, San Diego, CA, United States
¹¹Department of Psychiatry, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
¹²Department of Psychiatry, Seoul National University College of Medicine, Seoul, Republic of Korea
* these authors contributed equally

Corresponding Author:
Jennifer G Kim, PhD
Georgia Institute of Technology
225 North Ave
Atlanta, GA, 30332
United States
Phone: 1 404 894 2000
Email: jennifer.kim@cc.gatech.edu

Abstract

Background: Individuals with autism often experience heightened anxiety in workplace environments because of challenges in communication and sensory overload. As these experiences can result in negative self-image, promoting their self-efficacy in the workplace is crucial. Virtual reality (VR) systems have emerged as promising tools for enhancing the self-efficacy of individuals with autism in navigating social scenarios, aiding in the identification of anxiety-inducing situations, and preparing for real-world interactions. However, there is limited research exploring the potential of VR to enhance self-efficacy by facilitating an understanding of emotional and physiological states during social skills practice.

Objective: This study aims to develop and evaluate a VR system that enabled users to experience simulated work-related social scenarios and reflect on their behavioral and physiological data through data visualizations. We intended to investigate how these data, combined with the simulations, can support individuals with autism in building their self-efficacy in social skills.

Methods: We developed WorkplaceVR, a comprehensive VR system designed for engagement in simulated work-related social scenarios, supplemented with data-driven reflections of users’ behavioral and physiological responses. A within-subject deployment study was subsequently conducted with 14 young adults with autism to examine WorkplaceVR’s feasibility. A mixed methods approach was used, compassing pre- and postsystem use assessments of participants’ self-efficacy perceptions.

Results: The study results revealed WorkplaceVR’s effectiveness in enhancing social skills and self-efficacy among individuals with autism. First, participants exhibited a statistically significant increase in perceived self-efficacy following their engagement with the VR system (P=.02). Second, thematic analysis of the interview data confirmed that the VR system and reflections on the data fostered increased self-awareness among participants about social situations that trigger their anxiety, as well as the
behaviors they exhibit during anxious moments. This increased self-awareness prompted the participants to recollect their related experiences in the real world and articulate anxiety management strategies. Furthermore, the insights uncovered motivated participants to engage in self-advocacy, as they wanted to share the insights with others.

Conclusions: This study highlights the potential of VR simulations enriched with physiological and behavioral sensing as a valuable tool for augmenting self-efficacy in workplace social interactions for individuals with autism. Data reflection facilitated by physiological sensors helped participants with autism become more self-aware of their emotions and behaviors, advocate for their characteristics, and develop positive self-beliefs.

(JMIR Form Res 2024;8:e52157) doi:10.2196/52157

KEYWORDS

autism; virtual reality; workplace; self-efficacy; social skills; data reflection

Introduction

Background

Approximately 40% of individuals with autism experience anxiety because of difficulties in socializing, sensory sensitivities, and other factors [1-3]. Specifically, workplace environments can amplify anxiety in individuals with autism, such as sensory overload, communication barriers, and unplanned interactions [4-8] as these environments are often designed with neurotypical expectations in mind. When people with autism constantly encounter situations that elicit anxiety, it can cause them to have negative self-beliefs about their skills and performance, which can lower their self-confidence in the workplace [9]. Therefore, self-efficacy—the personal judgment or belief in one’s ability to succeed in prospective situations—is crucial for people with autism because it can help individuals approach challenging workplace experiences from a positive perspective, as opposed to focusing on failure [10,11].

Virtual reality (VR) interventions have demonstrated potential effectiveness in promoting the self-efficacy of people with autism, as they provide a safe yet realistic environment to master specific social skills by offering opportunities to repetitively practice them [12-18]. Although this exposure and repetition are useful for mastering skills, gaining a deeper understanding of specific social situations that trigger anxiety can better prepare users to effectively manage those situations in the real world. An individual’s understanding of anxiety-inducing situations and crafting self-beliefs that they have the capabilities to succeed in those situations align with 2 key constructs of the self-efficacy theory by Bandura [19,20]—physiological states and verbal persuasion. Notably, VR can also support physiological state awareness and verbal persuasion. It can identify the VR situations that prompt significant changes in a user’s physiological data, enabling the user to reflect on these experiences and formulate strategies to respond effectively [18]. This process can empower the development of self-efficacy [19]. However, existing VR interventions often limit the scope of self-efficacy to mastery experiences only [21] and fail to provide the extended theoretical rationale or background of self-efficacy on the design or outcomes of the system.

In this study, we expanded the current VR system by incorporating a feedback model in which the user’s behavioral and physiological sensor data can be reviewed immediately following the experience simulation, allowing for situational reflection.

Our research aimed to increase self-efficacy through a VR-based social skill training system for individuals with autism. Furthermore, we investigated how the facilitation of self-understanding through incorporating reflection of physiological and behavioral data immediately after the social skills practice can impact self-efficacy in preparation for real-life scenarios.

Objectives

This study had 2 main objectives. First, we developed WorkplaceVR, a VR application that allows people with autism to engage in simulated work-related scenarios to help understand their performance through data-driven reflection, with users’ behavioral and physiological data collected during the VR experience. Second, we evaluated the feasibility of the WorkplaceVR by conducting a deployment study with 14 young adults with autism. Using a mixed methods approach, we investigated the changes in self-efficacy among users with autism using pre- and postsurvey questionnaires. In addition, we conducted interviews to identify how participants with autism use, expect, and encounter challenges in the VR experience.

Methods

In this study, we designed and implemented WorkplaceVR, a VR-based system that offers simulations of work-related social situations and data-driven reflection of users’ behavioral and physiological responses (Figure 1).
Figure 1. WorkplaceVR is a virtual reality–based system that offers immersive experiences of work-related social situations. With WorkplaceVR, our participants practiced their social skills in a simulated café environment in 2 basic-level and 2 advanced-level scenarios. Our findings highlighted that participants with autism were actively engaged in WorkplaceVR: placing their hands on the physical desk where the virtual counter was (top) and turning to face the virtual manager when conversing (bottom).

Phase 1: Development of the VR System (WorkplaceVR)

Inclusive and Iterative Design for the VR Program Development

Our user-centered design approach to developing the VR system is to ensure inclusivity; technology must empower a more diverse and inclusive society [22,23]. We aimed to build a safe and less stressful virtual environment for people with autism, where they can become familiar with the workplace environment and practice interacting with others without fear of failure. Moreover, with inclusive design in mind, we conducted an iterative design process to incorporate the voices of people with autism into the VR program.

We first conducted a preliminary study to draw design suggestions for the VR program [24]. To elicit end-user feedback, we created a 5-minute video prototype demonstrating the overarching concept and the use scenario of the WorkplaceVR program. Using the video prototype, we conducted semistructured interviews with 20 participants, including individuals with autism (employees, job seekers, etc), managers of companies where people with autism are currently working, psychiatrists, and professional counselors, to uncover the various needs of individuals with autism at workplaces. Through the interview results, we identified the following key design insights. First, the participants emphasized the importance of designing realistic VR scenarios to help users engage in the program by reflecting on their personal experiences or challenges at work. Second, the system should guide users to reflect on their emotions and behaviors. Finally, during the data reflection phase, participants with autism should be able to take control of expressing their thoughts and behaviors and build confidence.

With insights from the preliminary study, we developed our WorkplaceVR program (Figure 1). The system comprises 2 parts: the simulation VR scenario and the physiological data visualization interface. WorkplaceVR was developed using the Unity3D engine (Unity Technologies) and the SteamVR plug-in (Valve Inc) and runs on a Windows 10 (Microsoft) PC with an Intel Core i7, GeForce RTX 2070 graphics card, and 16 GB RAM. A head-mounted display (HMD), VIVE Pro Eye VR headset (HTC) [25], and Empatica E4 wristband (E4 band, Empatica Inc) [22] were worn by the users for viewing the virtual world and sensing physiological signals.

We conducted a pilot study with 4 neurotypical participants to assess the study protocol and identify any risks and challenges that users might experience in trying WorkplaceVR. The participants were introduced to the VR system, experienced WorkplaceVR using an HMD, and then asked to provide feedback on the overall experiences, including task difficulty and scenario length, visual components of the interface, and side effects of the VR experience if there were any (eg, motion sickness, headache, and visual disturbance). On the basis of participants’ suggestions, we added additional visual cues, such as a shining effect to objects (eg, the bill receipt) for users to quickly find and interact with in the 3D environment and an arrow user interaction icon to lead the eye to a specific direction or object to draw attention to specific objects or areas of interest when needed. Another problem raised by the users was regarding the test environment setup. To facilitate an environment where the users can feel safe and private, we placed a blackout curtain in front of the desk where they stood during the experiment.

WorkplaceVR: System Design

In this section, we elaborate on the features and design considerations of the VR system, WorkplaceVR, based on the self-efficacy theory by Bandura [19,20].

Designing VR Contents for Mastery Experience

VR can offer mastery experiences by simulating real-world scenarios that can engage users with autism to accomplish tasks related to workplace settings. To achieve this, our design
approach focused on three key factors: (1) the inclusion of authentic work scenarios, (2) the integration of evidence-based social skills, and (3) the incorporation of multiple levels of difficulty within the scenarios.

1. Realistic work scenarios: we used café scenarios with work-related realistic visuals and auditory stimuli that provide a sense of surrounding in the immersive environment. For example, we designed café furniture and appliances (e.g., a cash machine and a coffee machine) and instrumental background music to create a relaxing café atmosphere. Specifically, we selected a situation in which the barista must inform a customer that the drink they ordered is ready from among the situations that are required when working in a café (e.g., greeting customers upon entry and taking customer drink orders).

2. Evidence-based social skills intervention: in specific café-based scenarios involving interactions between a barista and a customer, we incorporated evidence-based social skills (e.g., active listening, initiating conversations, and not interrupting when someone else is talking) derived from interventions supported by previous research [26,27]. In addition, we included context-specific skills (such as informing customers about available options, verifying order accuracy, and problem-solving in unforeseen circumstances) sourced from the café service training manual [28,29].

3. Levels of scenarios: for the users to build self-efficacy, it is important to help them reduce anxiety and fear of failure by designing attainable goals that gradually increase in difficulty [19,30-32]. Thus, we designed 4 scenarios with 2 levels of difficulty: basic and advanced. The basic level requires the users to explore and interact with a simulated 3D environment to practice basic conversational skills as baristas. Following this, the users are involved in starting a conversation with a customer avatar by offering a polite greeting, taking orders, and serving beverages such as coffee. The 2 scenarios at this level have the same structure but different order details (number and types of beverages). When the users perform these tasks successfully, they are moved on to the advanced-level scenarios. At the advanced level, the scenario adds complications that arise within the conversation with the customer avatar (e.g., a customer avatar claiming that their drink order is wrong).

Physiological Data and Visualization Interface

In the self-efficacy theory, Bandura [19,20] explains that people can shape their perceptions and beliefs about their capabilities by examining one’s physiological and emotional states. Therefore, one of our goals in designing the VR program was to help users better understand their emotions and behaviors by identifying their physiological states and behaviors in social situations [12,33]. We aimed to support people with autism to become more aware of their physiological and emotional states when they face stressful or anxiety-raising situations in the virtual work environment. Hence, we collected the users’ behavioral and physiological data during the VR intervention using the E4 band and HMD. We synchronized the time stamp information transmitted from E4 with the HMD time stamp information using the open-source framework Flask [34]. In addition, we presented a data visualization interface for users to understand and interpret their data along with their recorded performance videos, as illustrated in Figure 2. On the interface, we present (1) anxiety-related physiological measures (e.g., temperature, electrodermal activity [EDA], and heart rate [HR]; Figure 2B); (2) changes in voice volume (Figure 2C); and (3) detection of eye contact (Figure 2D). The following explains how we defined, processed, and visualized each data type: anxiety moments, voice volume, and eye contact.

1. Anxiety moments: anxiety moments were defined using the time stamp on data collected when the sensors detected radical changes in the signal because of the users’ physiological reactions to the stimuli. We calculated anxiety moments using an anomaly detector provided by the Microsoft Azure machine learning algorithm [35]. This involved 5 sensor types—HR, EDA, temperature, interbeat intervals, and blood volume pulse—provided by the E4 band. We presented the anxiety moments on the interface with a time stamp range in seconds; each time stamp is linked to the exact time of the performance video. Users can click on a time range to watch their performance video on the left, as shown in Figure 2A.

2. Voice volume: voice data were collected using the microphone of the HMD. We then presented the voice volume level as a graph with time stamps to help users recognize changes in their voice tone and volume (Figure 2C).

3. Eye contact: to help participants understand their eye movements during their moments of anxiety, we used a region of interest (ROI), which is a specific area within an image or video selected for analysis. In our study, we set the faces of the customer and manager avatars as ROIs. We used the box collider of Unity3D and measured whether a user gazes at ROIs through Tobii G2oM [36], which is a machine learning algorithm that can accurately ascertain on which objects a user focuses. If the participant looked anywhere in that location, it was labeled as seen (1); if they looked at another area, it was labeled not seen (0) in 1-second units. On the basis of the labeling results, we presented the seen labeled periods as a bar graph to help users understand how they make eye contact with people during VR scenarios, as presented in Figure 2D.
Verbal Persuasion Through User-Driven Insights
In this study, we focused on facilitating a data-driven reflection process to leverage verbal persuasion for users with autism to engage in the sense making of their physiological data. Our goal was to provide an opportunity for users with autism to navigate their physiological and behavioral data results, identify their own strengths and interests, and gain confidence in speaking about themselves. Thus, the participants were given an active role in interpreting their data and answering questions such as What insights have you gained from the data about your characteristics or strengths? and What are your goals, taking data into account?

Phase 2: Implementation of the VR Program With Participants With Autism
We recruited participants with autism aged >16 years who are either currently employed or unemployed but plan on job searching in the near future. Our inclusion criteria for participants with autism were people who (1) have been diagnosed with autism; (2) can verbally articulate their thoughts, feelings, and experiences; (3) have no difficulties wearing an HMD for VR (eg, sickness or headaches while in VR or vision impairment, such as anisometropia); and (4) understand the study procedure and agree to participate. We posted the study posters to autism-related web-based communities on social networking services and web-based bulletin boards of autism-related institutions (eg, developmental disability social welfare centers and employment agencies for people with disability). In addition, we placed the flyers on the offline bulletin board of a child and adolescent psychiatry outpatient clinic in a hospital and a private counseling center for individuals with autism to outline the participant demographics.

Ethical Considerations
To ensure ethical conduct, our study received approval from the Institutional Review Board of Seoul National University Bundang Hospital (institutional review board number: B-2202-736-302). As this study involved the collection of sensitive information (eg, physiological data) from participants, we informed the participants about the data collection process and obtained their informed consent before they participated in the study. The researchers explained the consent form to the participants with autism in easily understandable terms. We also clarified to the participants with autism that any personal information, VR data and sensor data collected during the research would be anonymized for analysis and securely discarded to protect their privacy. In addition, we offered a compensation of US $50 for the research participants.

Measures
Perceived Self-Efficacy for VR Social Skill Training Scale
We developed the Perceived Self-Efficacy for VR Social Skill Training Scale (PSES-VR), an 8-item questionnaire with a 5-point Likert scale, to evaluate whether our VR intervention changed people’s beliefs regarding the self-efficacy of practicing social skills at the workplace. We developed this scale by modifying the Perceived Self-Efficacy questionnaire based on the theory of perceived self-efficacy by Bandura [37] and referencing existing scales, including the Perceived Social Self-Efficacy Scale [37], Perceived Improvement Scale [38], and Self-Efficacy Scale for Social Workers [39]. The questionnaires were revised to address the target social skills in the VR scenario. The 6 items consisted of communication skills required in general conversation situations. The 2 items evaluate whether participants respond appropriately to the situation required in the VR program scenario. Higher total scores on the items indicate better self-efficacy related to the social skills of the participants. The participants were asked to complete the same PSES-VR survey before the user study session and after the VR experience. All items of the PSES-VR are attached in Multimedia Appendix 1.
**iGroup Presence Questionnaire**

We used the iGroup Presence Questionnaire (IPQ), a 14-item questionnaire with a 7-point Likert scale, to investigate how users perceive a sense of presence of our VR system. The IPQ scale consists of 4 components: a general sense of being there (1 question), the sense of spatial presence (5 questions), involvement (4 questions), and experienced realism (4 questions), measuring the level of perceived presence during the VR experience [40]. Higher scores on the 4 components, as well as the total scores, indicate a better sense of presence in the VR system as perceived by the participants. All items of the IPQ are attached in Multimedia Appendix 1.

**Study Procedure**

Our study procedure included three stages: (1) pre-experiment, (2) VR experiment, and (3) postexperiment reflection phase. Figure 3 presents a summary of the procedure.

**Before the Experiment**

To start, the researchers provided a brief overview of the study purpose and conducted a brief interview with participants, asking questions about their work-related experiences and previous experiences with VR. Next, the researchers introduced WorkplaceVR with explanations of the contents of the VR system and sensor devices that the participants should wear during the VR experiment. After that, the participants had a tutorial session to learn the 2 basic-level scenarios through video and scripted role-play.

**VR Experiment**

When the participants were ready for the VR experiment, they were asked to wear the E4 band, which is used for sensing the physiological signals, and to put on the VR headset. We informed the participants that they could stop at any time if they started experiencing motion sickness. After they wore the devices, we asked the participants to explore the café environment to adjust to the VR environment. Then, all participants performed the same 4 scenarios in the order of basic to advanced scenarios in VR for approximately 10 to 15 minutes.

**After the Experiment**

After the VR experience, we took a 5-minute break and conducted a data reflection on their VR performance with sensor data. In the data reflection process, participants were presented with various types of data related to their own anxiety moments, voice volume level, and eye contact while video recording their performance in VR using the interfaces and were asked to reflect on their performance based on these data. Here, we informed the participants that the data may not be accurate because of technical issues, so there can be errors or missing data in the interface. In the data reflection process, we asked the following to help participants reflect their data, such as “How was your performance in the VR scenarios?” “What were your strengths and challenges while performing the VR scenarios?” or “Did you find anything new about your behaviors while reviewing the data?” After the data reflection, we conducted a semistructured interview about the overall VR experience and data reflection. We asked participants how the VR content (eg, tasks and levels of scenarios, VR environment, and avatar design), data-driven reflection, and protocol (eg, watching a video and role-playing before the VR experiment) could be used to better understand their emotions and responses in potential workplace settings. We also asked questions to elicit feedback on the usefulness of data-driven reflection and its potential impact on their self-understanding of their behavior.

**Data Analysis**

For qualitative data analysis, the interview data were transcribed, coded, and analyzed based on open coding and thematic analysis [41] using ATLAS.ti (version 7; Scientific Software Development GmbH). A total of 3 researchers individually read the interview transcripts and generated open codes that were discussed among the research team to identify patterns and build themes around VR experience and data-driven reflection regarding self-efficacy. The coding procedure was iterative; it concluded once the researchers agreed that the themes were consistent and a distinct set of themes surfaced. Interviews and surveys were conducted and documented in Korean, and analyses were performed in the same language. This ensured that the original responses were analyzed with the utmost precision. For the quotations used in the article, an English-Korean bilingual-speaking researcher initially translated the responses into English and then revised them through consultation with a proofreading expert.

For quantitative analysis, we analyzed the scale results (IPQ and PSES-VR) of the participants with autism. As a case-control
group was not included in this study, it was difficult to confirm the statistical significance of the IPQ scale in identifying the presence of the VR system in participants with autism. For the IPQ scale, we calculated the mean and SD of the items corresponding to each of the 4 IPQ components and used it descriptively to analyze the VR presence of participants with autism. For the PSES-VR, we performed a paired 2-tailed t test to analyze whether there was a statistically significant change in perceived self-efficacy after using our VR system. For anxiety moments and eye contact sensor data, we conducted a 1-way ANOVA to analyze whether there was a statistically significant difference among the 4 scenarios in WorkplaceVR.

### Results

#### Overview of the VR Experience

A total of individuals with autism participated in the study, including 2 women and 12 men, with an age range of 16 to 34 years. Table 1 shows the baseline characteristics of the participants.

There were no participants who reported difficulties such as motion sickness and headache during the VR experiments. Most (12/14, 86%) participants succeeded in completing the advanced-level scenarios without any support. In total, 2 participants (ND8 and ND13) completed the advanced-level scenario with minimal prompts from the researchers, such as guiding participants to find where the receipt was placed in the table.

#### Table 1. Demographic information about participants.

<table>
<thead>
<tr>
<th>Code</th>
<th>Sex</th>
<th>Age (y)</th>
<th>Work experience (work period)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ND1</td>
<td>Male</td>
<td>23</td>
<td>Food service experience at a fast-food restaurant (1 y)</td>
</tr>
<tr>
<td>ND2</td>
<td>Male</td>
<td>32</td>
<td>Office worker (8 y)</td>
</tr>
<tr>
<td>ND3</td>
<td>Male</td>
<td>23</td>
<td>Undergraduate student or part-time job (assistant at a counseling center; 2 y)</td>
</tr>
<tr>
<td>ND4</td>
<td>Male</td>
<td>24</td>
<td>Part-time for a cleaning service and daily paid jobs (event staff; 4 y)</td>
</tr>
<tr>
<td>ND5</td>
<td>Female</td>
<td>22</td>
<td>Undergraduate student or daily paid jobs (PowerPoint presentation design; 1 mo)</td>
</tr>
<tr>
<td>ND6</td>
<td>Male</td>
<td>19</td>
<td>Undergraduate student and no working experience</td>
</tr>
<tr>
<td>ND7</td>
<td>Male</td>
<td>27</td>
<td>Undergraduate student or café barista (1 y)</td>
</tr>
<tr>
<td>ND8</td>
<td>Male</td>
<td>25</td>
<td>Remote worker (data entry in Excel) or designer (making web-based banners; 3 y)</td>
</tr>
<tr>
<td>ND9</td>
<td>Male</td>
<td>20</td>
<td>Undergraduate student and no working experience</td>
</tr>
<tr>
<td>ND10</td>
<td>Male</td>
<td>20</td>
<td>Undergraduate student or part-time work experience (warehouse loading truck job and restaurant server; 3 mo)</td>
</tr>
<tr>
<td>ND11</td>
<td>Male</td>
<td>20</td>
<td>Customer service agent (8 mo); staff at a central radio management service (4 mo)</td>
</tr>
<tr>
<td>ND12</td>
<td>Female</td>
<td>27</td>
<td>Part-time worker (pet care; 1 y)</td>
</tr>
<tr>
<td>ND13</td>
<td>Male</td>
<td>33</td>
<td>Cleaning and maintenance (2 y)</td>
</tr>
<tr>
<td>ND14</td>
<td>Male</td>
<td>21</td>
<td>Freelancer (6 mo)</td>
</tr>
</tbody>
</table>

#### Quantitative Assessment of the VR System Use

Our results showed a significant increase in the perceived self-efficacy of participants with autism ($P=.02$) before and after experiencing WorkplaceVR. In the IPQ result, the mean scores for “general presence” (or the “sense of being there”) and “experienced realism” were higher than the mean scores for the other 2 components (spatial presence and involvement). Table 2 presents the results of the questionnaires.

Table 3 presents the physiological sensor data of the participants in the 4 VR scenarios.

Our hypothesis was that participants with autism would experience more anxiety moments in the advanced scenarios. Consistent with the hypothesis, the participants had more anxiety moments in the advanced scenarios than in the basic scenarios ($P<.001$). This result shows that our physiological sensing data can effectively act as a proxy indicator for anxiety experienced by participants with autism. This is because individuals with autism tend to encounter elevated levels of anxiety in unexpected situations, such as the more advanced scenarios that we introduced.
Table 2. Results of the Self-Efficacy Scale and System Evaluation.

<table>
<thead>
<tr>
<th>Scales</th>
<th>Values, mean (SD)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSES-VRa</td>
<td></td>
<td>.02b</td>
</tr>
<tr>
<td>Pre</td>
<td>21.86 (7.33)</td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>39.07 (7.52)</td>
<td></td>
</tr>
<tr>
<td>IPQc</td>
<td></td>
<td>N/A d</td>
</tr>
<tr>
<td>General presence</td>
<td>4.07 (1.86)</td>
<td></td>
</tr>
<tr>
<td>Spatial presence</td>
<td>3.87 (0.59)</td>
<td></td>
</tr>
<tr>
<td>Involvement</td>
<td>3.71 (0.79)</td>
<td></td>
</tr>
<tr>
<td>Experienced realism</td>
<td>3.89 (1.34)</td>
<td></td>
</tr>
</tbody>
</table>

aPSES-VR: Perceived Self-Efficacy for VR Social Skill Training Scale.
bThere was a significant increase in the perceived self-efficacy of participants with autism before and after their experience with WorkplaceVR.
cIPQ: iGroup Presence Questionnaire.
dN/A: not applicable.

Table 3. Physiological data of participants at each scenario.

<table>
<thead>
<tr>
<th>Sensor data</th>
<th>Basic scenarios</th>
<th>Advanced scenarios</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First</td>
<td>Second</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>First</td>
<td>Second</td>
</tr>
<tr>
<td>Anxiety moments&lt;sup&gt;a&lt;/sup&gt; (number of points), mean (SD)</td>
<td>2.43 (3.56)</td>
<td>3.21 (3.36)</td>
<td>9.43 (4.96)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Anxiety moments are defined as the timestamps detected by the Azure algorithm’s anomaly detector when sensor signals (eg, electrodermal activity and heart rate) display fluctuations in response to users’ physiological reactions to stimuli.

<sup>b</sup>There were significantly more anxiety moments in the advanced scenarios than the basic scenarios.

Qualitative Analysis of the VR System’s Impact on the Self-Efficacy

In this study, qualitative data analysis revealed 3 major themes that correspond to 3 of the 4 constructs of the self-efficacy theory by Bandura: physiological responses, verbal persuasion considered as encouragements, and mastery experience [19]. Below, we report how these findings can shed light on the feasibility of using the WorkplaceVR system to promote self-efficacy.

Impact of the VR Program on Self-Efficacy of Participants With Autism

Self-Awareness of Physiological and Emotional State Through Data Reflection

In the data reflection process, we provided an interface for visualizing the sensor data (ie, anxiety moments, voice volume level, and eye contact) to help the participants better recognize their emotions and behaviors in relation to their VR experiences. All our participants reviewed whether the presented data accurately reflected their emotions and behaviors. While reviewing their physiological signals (eg, HR, IBI, and EDA) presented by the VR system, the participants explained what they were feeling, thinking, or doing. For example, ND6 noticed that his HR increased when he was experiencing tense situations in VR and reported that he was aware of the stressors and reactions, drawing from his prior experiences:

> I know that my heartbeat increases when I am in situations that make me anxious or nervous. I remember my heart beating so fast on the night of June 29th. There was also thunder that night [...]. It was like the nervous feeling I get when I check my grades. [ND6]

Moreover, the participants recognized and described the patterns of when and why they felt anxious in certain situations. By watching the recorded video of their performance at “anxiety moments,” participants actively described when they were most stressed or anxious while carrying out the simulated workplace tasks (eg, when a customer avatar approaches or when a customer avatar makes a sudden request to the user). For instance, ND2 had 3 events marked as “anxiety moments,” all of which were related to tasks where the user was required to start a conversation with the customer avatar. After playing back the recorded videos for all 3 events, ND2 reported that the data well represented his characters and that he usually becomes anxious when he has to initiate conversations with people at work:

> Oh, I’m quite sensitive sometimes, especially before I start talking, because I feel like I have to say...
By reflecting on their physiological and emotional states and behaviors in advanced scenarios, it enabled participants to recollect their related experiences in the real world and talk about goals to tame anxiety. For example, after reflecting on anxious moments, ND6 discussed ways in which he can improve his comfort in similar situations in the future:

I don’t want to panic again when a problem arises. For example, I shouldn’t be too stressed about my exam results. Whether it be a good or bad result, what’s important is accepting the result and looking beyond it. [ND6]

In addition, some participants suggested that having more diverse data (eg, facial expressions, body movements, and gestures) could help them better reflect on and understand their behaviors:

It would be helpful to see how I make gestures and move my body because it is something that is hard to notice in real life. Having an observation camera might work. [ND7]

Verbal Persuasion Through Self-Expression in Data Reflection

We found that the data-driven reflection process of our system can help participants with autism better understand themselves and build positive self-beliefs about who they are and what they can achieve. This opportunity allowed participants with autism to engage in verbal persuasion experiences of self-efficacy [42,43], which means that they could speak about their own strengths and interests and gain confidence in expressing themselves:

I was worried that I am not usually good at making eye contact. But seeing the data, I am surprised to realize that I am, in fact, doing quite well on this. I’m feeling more confident about myself. [ND10]

In addition, some participants reported that they wanted to use behavioral and physiological data to explain the strengths and characteristics they recently discovered to others so they could express their thoughts or arguments more clearly. For example, ND12 reported that she feels misunderstood by people when they have the wrong impressions or ideas about her. She proposed ways to use the data to effectively communicate her opinions to decrease potential conflicts with others:

I often find myself in a state of persistent frustration when communicating with others. I believe that people do not truly understand or accept my thoughts or feelings. [...] However, with the physiological sensor data I currently possess, I am confident that I can convince others. [ND12]

Mastery Experience Through Realistic VR Scenarios

Our interview results suggested that participants with autism experienced a sense of mastery while using the VR program. For example, the participants with autism reported that the successful completion of realistic workplace interpersonal situations in a VR environment boosted their confidence in handling similar situations in the real world:

You have to taste the fruits of success in order to gain confidence and not be afraid of failure. I felt this was an important factor for me to go on to the advanced level because failure can have a huge damaging effect. [ND4]

This mastery experience was also supported by our VR system design that gradually increases the level of task difficulty from the basic-level scenarios to help users respond to unexpected situations with confidence:

Although it was new, it didn’t deviate too much from what I already experienced, so I could handle it. [ND14]

Consistent with the IPQ result, the participants also reported a high level of presence in WorkplaceVR. The participants explained that this was because our VR program provided high-fidelity simulation embodying realistic visual content (eg, “Face Mask Required” signs on the walls of the virtual café considering COVID-19 pandemic situations) and immersive scenarios where they could have a naturalistic social interaction with virtual customers and colleagues (eg, using gestures to communicate and giving receipts to customers). For example, notable observations indicating high engagement included participants attempting tasks such as making coffee or using the cash machine, even though these actions were not part of the assigned tasks (Figure 1). The lifelike experiences participants experienced during our VR program could have played a role in cultivating a sense of mastery.

Discussion

Principal Findings

Overview

This study demonstrated that the VR program, which enables individuals with autism to experience work-related social scenarios and reflect on their VR practice through physiological and behavioral data visualization, can significantly increase the individuals’ perceived self-efficacy in practicing social skills within a workplace context. The interview data further showed how the data reflection of VR practice can influence self-efficacy. The user-driven data review practice allowed individuals with autism to reflect on their physiological data, that is, by promoting self-awareness of their emotions, gaining insights into their real-world behaviors that they were unaware of, and self-advocating their characteristics to others based on their data. In particular, participants could understand when and why they feel anxious, enabling them to proactively devise strategies for self-comfort in anticipation of similar anxiety-provoking situations. Moreover, the increased self-awareness about the underlying causes of their anxiety and
related behaviors motivated them to communicate their experiences and advocate for their needs with others.

**Promoting Self-Efficacy by Promoting Self-Awareness About Physiological States**

We found that the physiological data reflection helped individuals better understand their emotional responses. This increased self-awareness that participants with autism obtained through our VR system motivated them to take their learnings from reflection in the real world to better explain and advocate for themselves to others. Therefore, in this section, we discuss how the self-reflection interface should be designed to present physiological and behavioral data in a way that encourages individuals to reflect on their experiences.

According to Bandura [19], recognizing and managing one’s own emotions and physiological states is essential in promoting self-efficacy, as this affects people’s decision-making process and performance. In line with the theory, our study found that behavioral and physiological sensor data (eg, eye contact, voice volume, EDA, and HR) could be used to support people with autism to become aware of their emotional states and, in turn, establish strategies to respond effectively to intense emotions. The participants became aware of their current affect state by mapping their physiological responses on the interface and sometimes wanted to examine data in depth to improve their self-understanding (eg, facial expression, standing posture, and hand gestures). Specifically, reflecting on the physiological data taken during the VR experience while watching the playback of the sessions allowed participants with autism to revisit how they felt and behaved in VR situations in specific moments that heightened their anxiety. This reflection reminded them of similar situations that induce anxiety in their daily lives, such as when they had to initiate conversations or when a conversational partner is approaching them. This increased self-awareness about their emotional responses—why they were anxious or nervous at specific times—further allowed participants with autism to make a resolution: how they might manage their emotional reactions in everyday lives.

Moreover, we found that the data reflection allowed participants with autism to gain insights about their real-world behaviors that they were unaware of (eg, “Why I was anxious when I had a conversation with coworkers?” or “Why I was always tired when I went to a place with a lot of people?”). Therefore, VR interventions for promoting self-efficacy could be designed to provide opportunities for users to investigate their emotional reactions through data reflection and to connect the insights to their real-world practices. To encourage reflection, Fleck and Fitzpatrick [44] suggest incorporating reflective questions into technology to prompt users to think about their behaviors and provide justifications or explanations for knowledge, actions, or events. Therefore, VR systems can present reflective questions such as “What about this scenario made you feel anxious?” to accompany data reflection and to provide scaffolding for individuals with autism to consider how their performance in the scenarios relates to their real-world experiences.

Finally, when reflecting on the physiological data marked during the VR simulations, the system can guide users to raise their emotional awareness by relating it to their real-world experiences. It is important for them to understand what factors trigger their emotional and physical responses, why these factors affect them in a certain way, and how they should respond to such emotions. Previous research suggests that careful observation of one’s behavior, either by themselves or by others, might be the most informative and applicable source of emotional self-awareness [45,46]. Similarly, in our study, participants with autism identified the situations where they felt anxious while reviewing the anxiety moments data and watching the recorded videos of their VR performance and described their feelings by recalling prior related experiences. Through this process, they planned more specific ways to respond to anxiety, which could be applicable to their real-world interpersonal situations.

**Toward Data-Driven Self-Advocacy**

Verbal persuasion, involving encouragement from others and self-advocacy practice, is an important source of self-efficacy because it can help individuals shape self-beliefs that they have the skills or knowledge to succeed in a given situation and have confidence in themselves [19,47].

Our study suggests that the data-driven reflection process could have a similar effect as verbal persuasion. Our participants with autism reported that they often received negative feedback about their behaviors from others but experienced validation of behaviors through our system, for example, when the data interface indicated that their voice volume was lower than they expected. These results show that data reflection creates an opportunity for individuals with autism to experience verbal persuasion through identifying what they can do well in VR scenarios and fostering internal motivation to apply what they learned about themselves in the real world.

Furthermore, the participants with autism wanted to advocate for their characteristics or strengths identified in the data reflection to others to resolve conflicts or difficulties in their interpersonal relationships (eg, family members’ negative comments on the behaviors of autistic individuals). This finding suggests that data reflection could help individuals with autism to advocate for themselves in their daily lives and workplace. To design systems that can effectively support self-advocacy, our findings suggest the importance of presenting data relevant to their daily lives and supporting them to use the data to reflect on their behaviors, build confidence, and foster self-advocacy.

In our study, reflective questions [48] enabled users with autism to take time to understand themselves and translate their thoughts and concerns into positive resolutions based on the data. The questions included the following: “[reviewing the anxiety moment data] Have you ever encountered a similar situation in real life? If so, why did you feel that way? How do you typically respond to stressful events that make you anxious?” Although the participants could not directly manipulate the data visualization interface, future studies are needed to uncover how the interface can be designed to engage users with autism to better reflect on their personal interests, skills, and experiences. This could be approached by visualizing data with metaphors familiar to individuals with autism and
customizing the user interface to reflect users’ priorities and topics of interest [49-51].

Limitations
Although our VR system could provide individuals with autism with opportunities to promote self-efficacy, there are several limitations. First, in our study, we only included participants with autism who are able to communicate and interact with others. This decision was made in our study because WorkplaceVR was designed to focus on a specific population with autism. However, future research should explore how VR interventions can also benefit participants with autism who have different communication abilities. To estimate anxiety moments, we used the algorithm offered by Microsoft Azure. Although our participants confirmed that the anxiety predictions were aligned with their subjective feelings (e.g., anxiety and nervousness), future studies should investigate and apply more rigorous algorithms that can predict the anxiety levels of participants. In this study, we used the PSES-VR, a questionnaire written in Korean, and the IPQ, which was translated into Korean. However, neither of the 2 measures had been previously validated in the Korean population. Finally, we could not confirm whether participants’ experiences in the study would translate to their real-life situations through poststudy observations.

Conclusions
This study investigated how the VR system promotes the self-efficacy of individuals with autism for their success at work. For this, we presented WorkplaceVR, a VR system that allows users to experience work-related social situations and data reflection of the user’s behavioral and physiological data. Through the VR experiment and data reflection, we confirmed that the VR system significantly improved the perceived self-efficacy of participants with autism. Our study results revealed that the VR system provided participants with autism with an opportunity to have a mastery experience in VR scenarios, self-awareness of their emotional states, and self-advocacy of their strengths and characteristics through data reflection. By addressing the expectations and challenges in the VR system for people with autism, these results contribute to not only supporting researchers who design the technology for autistic people but also helping individuals with autism have a successful work experience.

Acknowledgments
The authors used generative artificial intelligence, ChatGPT (OpenAI, version 3.5, 2023), to correct the grammatical errors in the manuscript.

Data Availability
The data sets generated during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Contents of the two scales for evaluating social skills related to the perceived self-efficacy of the participants and the perceived sense of presence of our virtual reality system: (1) Perceived Self-Efficacy for VR Social Skill Training Scale and (2) iGroup Presence Questionnaire.

References

https://iformative.jmir.org/2024/1/e52157

JMIFFormRes2024 | vol. 8 | e52157 | p.134

(page number not for citation purposes)


25. VIVE Pro 2. VIVE. URL: https://www.vive.com/us/product/vive-pro-eye/overview/ [accessed 2023-08-16]


29. Service and training manual. Smith Coffee and Cafe. 2019 Jan 30. URL: https://static1.squarespace.com/static/5ad7d558f93fd4c94252090/t/5c54d0e4fd1046d014a4772292/1549062373251/SmithTraining.pdf [accessed 2023-12-22]


36. Tobii G2oM. Tobii. URL: https://vr.tobii.com/sdk/solutions/tobii-g2om/ [accessed 2023-08-16]


**PSES-VR**: Perceived Self-Efficacy for VR Social Skill Training Scale

**ROI**: region of interest

**VR**: virtual reality

Edited by A Mavragani; submitted 05.09.23; peer-reviewed by T Gargot, B Li; comments to author 20.10.23; revised version received 01.11.23; accepted 01.11.23; published 11.01.24.

Please cite as:


Promoting Self-Efficacy of Individuals With Autism in Practicing Social Skills in the Workplace Using Virtual Reality and Physiological Sensors: Mixed Methods Study

JMIR Form Res 2024;8:e52157

URL: https://formative.jmir.org/2024/1/e52157
doi:10.2196/52157

PMID:38206652

©Sung-In Kim, So-youn Jang, Taewan Kim, Bogoan Kim, Dayoung Jeong, Taehyung Noh, Mingon Jeong, Kaely Hall, Meelim Kim, Hee Jeong Yoo, Kyungsik Han, Hwajung Hong, Jennifer G Kim. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Mapping the Cardiometabolic Patient Experience and Self-Care Behaviors to Inform Design, Implementation, and Persistent Use of Digital Health Care Solutions: Mixed Methods Study

Jan Liska¹, MSc; Marie Mical¹, PharmD; Christophe Maillard¹, MA; Cécile Dessapt¹, CMPP, PhD; Europa Bendig², Dip Ing; Daniel Mai², MA, PhD; John D Piette³, MS, PhD; Sabina De Geest⁴,⁵, RN, PhD; Guillaume Fontaine⁶,⁷,⁸,⁹, RN, PhD

¹Sanofi, Paris, France
²STURM und DRANG, Hamburg, Germany
³Department of Health Behavior Health Education, School of Public Health, University of Michigan, Michigan, MI, United States
⁴Institute of Nursing Science, Department Public Health, University of Basel, Basel, Switzerland
⁵Academic Center for Nursing and Midwifery, Department of Public Health and Primary Care, Katholieke Universiteit Leuven, Leuven, Belgium
⁶Ingram School of Nursing, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada
⁷Centre for Clinical Epidemiology, Lady Davis Institute for Medical Research, Montreal, QC, Canada
⁸Centre for Nursing Research, Jewish General Hospital – Centre intégré universitaire de santé et de services sociaux West-Central, Montreal, QC, Canada
⁹Centre for Implementation Research, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, QC, Canada

Corresponding Author:
Jan Liska, MSc
Sanofi
46-48 Avenue de la Grande Armée
Paris, 75017
France
Phone: 33 141247000
Email: Jan.Liska@sanofi.com

Abstract

Background: Cardiometabolic conditions including acute coronary syndrome (ACS) and type 2 diabetes (T2D) require comprehensive care and patient engagement in self-care behaviors, and the drivers of those behaviors at the individual and health system level are still poorly understood.

Objective: We aim to gain insights into self-care behaviors of individuals with cardiometabolic conditions.

Methods: A convenience sample of 98 adult patients with ACS and T2D was recruited in the United States, Germany, and Taiwan to participate in a mixed methods study using ethnographic methods. All participants completed 7-day web-based diaries tracking their level of engagement, and 48 completed 90-minute web-based semistructured interviews between February 4, 2021, and March 27, 2021, focusing on themes including moments of engagement. Qualitative analysis identified factors influencing self-care practices and a Patient Mind States Model prototype.

Results: Patient reports indicate that many patients feel social pressure to adhere to treatment. Patients’ experience can be understood within 5 categories defined in terms of their degree of engagement and adherence (“ignoring,” “struggling,” “juggling,” “controlling,” and “reframing”).

Conclusions: For people living with ACS and T2D, the self-care journey is defined by patterns of patient experiences, which can identify areas that tailored digital health care interventions may play a meaningful role.

(JMIR Form Res 2024;8:e43683) doi:10.2196/43683

KEYWORDS
self-care; adherence; digital health; design; implementation; coronary; type 2 diabetes; care; patient engagement; behavior; interview; treatment; tool; digital tool; support
**Introduction**

The development of effective self-care behaviors is essential for patients living with chronic health conditions [1]. However, the definition of self-care behavior varies considerably between medical disciplines. From recent concept analyses, a broad term for self-care has been defined as “the ability to care for oneself through awareness, self-control, and self-reliance in order to achieve, maintain, or promote optimal health and well-being” [2]. Indeed, inadequate engagement in these self-care ability driving areas, suboptimal adherence to medication, and failure to enact healthy behaviors (eg, smoking cessation and physical activity) can compromise patients’ quality of life and health outcomes [1]. Self-care is particularly important for patients diagnosed with cardiometabolic conditions such as acute coronary syndrome (ACS) and type 2 diabetes (T2D) [3,4]. Despite this, rates of adherence to a care plan including but not limited to medication and lifestyle modifications remain suboptimal for this population [5-8], and a number of patient-specific factors (eg, depressive symptoms, attitudes toward management, and daily activities) that impact self-care and outcomes have been identified [9].

Despite efforts to establish the psychosocial factors underlying self-regulatory or self-management behaviors [10,11], the typical driver of such studies is to assess how these factors may impact adherence. Behavioral models have been considered within this research, such as the capability, opportunity, motivation, and behavioral model and intervention mapping; however, the volume of frameworks available and a fragmented, confusing taxonomy remain barriers to their effective use [12]. Equally, these frameworks are often generalized and applicable in a range of clinical scenarios and are not exclusive to diabetes or ACS.

This model design approach from a general perspective may explain in part why these models have had a limited impact in diabetes or cardiovascular care, where individual behavior plays a dominant role in patient outcomes [13]. Establishing a behavioral framework that is tailored for those with a cardiometabolic condition will form an important step toward understanding attitudes in clinical care and how these may innately fluctuate over time.

The determinants of effective self-care practices among people living with cardiometabolic diseases have been well-established and include self-efficacy, social support, cognitive skills, and positive attitudes [14-16]. Health care professionals (HCPs) and care systems play a central role in influencing self-care through these determinants as well as by optimizing health service delivery, addressing financial burdens, promoting engagement with technology, and encouraging community support initiatives [17,18]. As self-care determinants include both patient-level and system-level drivers, effective interventions must address both environmental factors and individual patient needs to achieve sustainable effects [19,20]. Designing these interventions requires a deep understanding of patients’ needs as well as their broader life context [21]. Patient activation–focused frameworks such as the Patient Activation Measure (PAM) or the Social, Psychological, Usage, Rational model have helped design interventions that improve medication adherence and lifestyle behaviors [22,23]. However, given ongoing changes in health systems’ use of digital patient support tools in the context of the COVID-19 pandemic, the ways in which tools such as the PAM and Social, Psychological, Usage, Rational model can best be used to support patient self-care needs further exploration. In particular, the COVID-19 pandemic has led to major shifts in clinical practice, including an increase in the use of digital health care approaches [24].

To better understand the ways in which health systems can support patients using digital support tools in the era of the pandemic, we conducted a survey of adult patients with ACS and T2D as well as their caregivers in the United States, Taiwan, and Germany. Our goal was to help establish a behavioral model based on the survey results, which would help inform how digital solutions may be utilized to improve adherence to therapy. This series of interviews and diaries, which drew on grounded theory and phenomenology, aimed to (1) provide deep human-centric insights into the behavioral dimension of the patient experience for people living with T2D or ACS and (2) identify the moments and motivational triggers in a patient’s life that have a strong impact on behavior and health outcomes.

**Methods**

**Research Design**

The survey consisted of 2 distinct phases, with all participants completing 7-day web-based, ethnographic diaries tracking their level of engagement in self-management and 48 participants completing web-based, semistructured interviews focused on themes including moments of engagement.

This survey used methods from the field of qualitative research, grounded theory [24] and phenomenology [25,26]. Grounded theory has been used to understand the processes through which patients manage new or chronic health problems [27], and as such, it has particular relevance to elucidating chronically ill patients’ experiences. Using grounded theory, data are collected and analyzed, and then a theory based on the resulting data was developed. The approach was designed to generate a theoretical explanation using both inductive and deductive approaches to a social phenomenon (ie, chronic disease self-care in this survey) from empirical data rather than a preconceived framework. Phenomenology is well suited to study the self-care of new or chronic health problems as it is based on the assumption that there is an essence to what people live with every day, and it aims to depict the basic structure of this experience [25]. As a qualitative research method that is particularly useful to study affective, emotional, and intense human experiences, phenomenology is a study of people’s conscious experience of their everyday life and social action [25].

**Ethical Considerations**

Informed consent was obtained from all individual participants included in this survey. The research is a qualitative behavioral research survey but not a clinical study or clinical survey. This research was conducted in accordance with the organization’s intended regulations for qualitative market research studies and no personal data or sensitive information were collected or presented in this publication. Information was recorded by the...
investigator in such a way that the identity of the human subjects cannot be directly or indirectly ascertained. This study was exempt from the institutional review board oversight in accordance with exemption guidelines listed in the 45 Code of Federal Regulations Part 46 and the Secretary’s Advisory Committee on Human Research Protections Recommendations on benign behavioral intervention of the Health and Human Services regulation in United States [28, 29]; the 2022 European Pharmaceutical Market Research Association Code of Conduct for the market research conducted in Germany [28]; and Article 5 of the Human Subjects Research Act of the Ministry of Health and Welfare, Republic of China (Taiwan), and the “Exempt Review Categories for Human Research” announced by Department of Health, Taiwan [29, 30].

**Sampling Methods**

Patients diagnosed with either ACS or T2D and cardiovascular comorbidities were identified and recruited in 3 countries: the United States, Germany, and Taiwan. These countries were selected based upon the Hofstede 6D model of cultural dimensions. The 3 countries were noted for their diversity of cultural dimensions that may influence and nurture patient behavior while holding similar maturity of health care systems.

We aimed to recruit equal proportions of men and women as well as patients with a variety in types of health insurance. Potential patients were members of an ongoing research panel who have agreed to be approached for research studies and have provided basic demographic and health status information. Patients with ACS and T2D were screened for eligibility using elements for ACS and T2D patient profiles, as well as survey-specific data regarding demographics, digital behavior, and personality profiles. As part of the screening process, patients were asked if they use any technology, either app or device, to manage their condition. All patients screened were interested and open to the idea of digital health care (ie, the use of digital tools to support tracking or managing measurements used to help manage their condition).

Screening was conducted by an external recruiting agency who preselected relevant candidates out of their patient panel. In total, 32 participants per country were selected out of approximately 48 recommended profiles per country from an external recruiting agency. An appropriate mix of people was selected by the research lead, striving for maximum variation in the sample among those who fulfill the sampling criteria.

Our goal was to recruit 16 patients with ACS from each country, including 8 patients who had undergone an invasive procedure (eg, percutaneous coronary intervention and coronary artery bypass graft). We also sought to recruit roughly equal numbers of patients with ACS who were aged 40-50 years and 50-70 years. All patients had to be eligible for enrollment in a cardiac rehabilitation program (whether this was digital or analog was recorded); approximately 60% (10/16) should have attended rehabilitation, while 40% (6/16) should not. A focus was placed on those with higher severity levels for their condition and those with progressed treatment regimens. Eligibility for enrollment in a cardiac rehabilitation program was determined as a proxy for ACS severity levels and an indicator of having undergone a medical procedure. In recruiting patients with ACS, our goal was to represent a diversity of family situations (single, married, and with or without children) and a range of experiences using self-care technology for managing cardiovascular conditions.

Finally, we sought to recruit 10-16 caregivers of patient with ACS to assist in setting up patient support with the diaries, 8 of whom would also serve as interview participants.

Our goal was to recruit 16 patients with T2D from each country, with equal numbers having less severe (eg, hypertension) and more severe (eg, heart arrhythmia and heart failure) comorbidities. These groups were each further equally split into those on oral antidiabetic drugs and those who had initiated basal insulin in the past 12 months (18-24 months if recruitment was difficult). All participants in the T2D sample were aged 40-70 years, with the group being split equally between those who currently use health care technology (an app or device) and those who are considered lapsed users of health technology. Within the screening process, patients with T2D were asked if they use any technology, either an app or device, to manage their disease. Lapsed users confirmed that they used technology to manage their condition but no longer do so.

Both the ACS and T2D samples were split into equal-sized groups of patients deemed to be fully adherent with medical advice and treatment and those considered either partly adherent or nonadherent. This was based on the assumption that the behavioral reasons for partial and nonadherence were similar. Adherence levels were determined based on patients’ self-reported assessment during the respondent recruitment process. Patients were asked a range of multiple-choice and open-ended questions, the results of which were used to quantitatively determine their adherence level. For instance, patients were asked to what extent they agree with the statement, “I am confident that I can follow through on medical recommendations my health care provider makes, such as changing my diet or regular exercise;” with which they could strongly agree, somewhat agree, somewhat disagree, or strongly disagree.

**Data Collection**

To increase internal validity of the survey through triangulation, 2 methods of data collection common in grounded theory and phenomenology research were used [26]. First, patients were asked to complete web-based diaries for 7 days, in which they recorded information including their daily experience of acute versus chronic; a mind map of engagement with notable life events, health, and conditions (eg, photos and video) for acute versus chronic; a mind map of helpers (people, tools, and institutions); and observations of self-concept versus reality (projections and narratives around the self-care experience).

The diaries consisted of 5 chapters, with each one completed once within 7 days. Each chapter was completed in consecutive order: (1) “Me and My Body,” (2) “My Story,” (3) “My Day-to-Day Life with My Condition,” (4) “My Daily Health Regimen,” and (5) “My Helpful Tools and Resources.” Daily diary exercise was recorded based on open-ended questions with private responses recorded via video, audio, photographs, a map, or written statement.
Specifically, respondents were asked to answer a range of open-ended questions over the course of 5 days. They were structured in six thematic buckets: (1) personal background, (2) me and my body, (3) daily health regimen, (4) day-to-day life with my condition, (5) helpful tools and resources, and (6) my health story. In topic 2, we asked, for instance, “Please recall particularly pleasurable and positive moments during the day when you felt happy, in control, proud of yourself, content with regards to a) general state of mind, b) relating to health, and c) relating to your condition.”

For the interviews, participants were screened by an external recruiting agency, relying on their patient panels, and an appropriate mix of people was selected by the research lead, striving for maximum variation in the sample among those who fulfill the sampling criteria. In the ACS sample, patients and their caregivers took part in dyad interviews. In the T2D sample, interviews were carried out on a one-to-one basis between interviewers and patients. These interviews aimed to capture key moments of engagement to gain an understanding of levers of change or opportunities; elucidate the role or influence of the caregiver; and explore a prototype of the patient experience map produced from patient diary content. The “SturmundDrang” team of cultural researchers and anthropologists created the interview guide in an iterative approach. This guide provided information on field research preparations. Designed as a rough framework of topics and themes to explore, this document offered guidance for the ethnographic video interviews. The same discussion guide was used across all countries, providing open-ended questions and sufficient room for exploration in different cultural contexts. The respective researcher could modify the guides in each country, if deemed necessary.

Data Analysis

Data were analyzed using a constant comparative method [31], which involves comparing 1 segment of data in an interview with another segment of data to determine similarities and differences [26]. The diary data was treated as 1 hermeneutic unit that was then qualitatively analyzed in direct comparison with the interview data. The interview data formed a second hermeneutic unit. Diaries and interviews were conducted within 2 weeks, allowing the researchers to pursue a close comparative approach. These 2 hermeneutic units were then compiled in a comprehensive raw field note document for each market, outlining the total qualitative data set that could then be analyzed.

Analysis was guided by the principles of horizontalization, where all data are treated with equal weight, and phenomenological reduction, which is the process of continually returning to the essence of the experience to derive the inner structure (of disease self-care) in and of itself [26]. We grouped data together on similar dimensions of (1) patient modes of engagement and experience domains, (2) the ongoing process of gaining self-care expertise, and (3) patient mind states regarding disease self-care. Furthermore, we identified beliefs and biases, drivers of engagement and challenges to disease self-care.

In particular, interviews were analyzed to identify emerging themes through established social-scientific methods of data gathering, including writing semistructured fieldnotes [32], qualitative data analysis [33], and ethnographic insights building [34]. To establish key patient mind states, the data were clustered using spatial clustering, iterative loops, and narrative listening. Spatial clustering of codes and signals was used to group themes into a figure of overarching mind states to provide a framework for qualitative interpretation and a hypothesis. Following data collection, the resulting map was refined through iterative loops, whereby the data and theory were examined and re-examined by a team of 3 researchers; with each loop, the number of clusters and mind states was adapted and refined. During the final iteration loop, the mind states and their names were co-designed with the patients to ensure a patient-centric outcome. Narrative listening was used throughout, with patients describing and naming the chapters of their journey to allow for patient-led clustering and refining of the mind states [35].

Results

Sample Characteristics

In total, 98 patients with ACS or T2D and cardiovascular comorbidities were recruited, 32 from the United States and 33 from each of Germany and Taiwan. All 98 patients who were recruited completed the survey. All patients completed the 7-day ethnographic diaries (Germany or United States: February 4-11, 2021; Taiwan: February 17-24, 2021). Patients spent approximately 5 hours to complete the research questions. In total, 48 web-based interviews were conducted (Germany or United States: March 3-15, 2021; Taiwan: March 22-27, 2021), 24 with patients with T2D and 24 with patients with ACS and their caregivers. Respondents were selected for further interviews based on the depth of patient journey detail.

Survey Results

Overview

From the survey conducted through interviews and diaries, a number of theoretical themes were identified.

Insights Into the Behavioral Dimension of the Patient Experience

Interviews with patients living with T2D indicated that the disease continues to extract a heavy emotional toll. Participants commented: “The nurse made me feel like some pig, that all I do is sit around and eat and be obese”; “I used to eat gourmet food anywhere, anytime, and now all I have left is pay attention to controlling my diet”; “If I were to buy something to eat – I almost never do – then I would have to take everything apart and weight the ingredients individually; I can then calculated what units I need to inject afterwards”; and “I thought I’d just take a pill and everything will be ok, so I didn’t take it as seriously as I should.”

These patient insights highlighted how T2D is still a chronic condition that complicates every aspect of one’s life. Overarching themes from these interviews indicate that the condition is widely perceived to be a self-inflicted lifestyle disease; patients experience feelings of stigma, shame, self-blame, and a need to justify their lifestyle choices to acquaintances and HCPs. The general experience is underscored
by patient sacrifices and unsolicited social pressures to adhere to treatment regimens; generating a depressive, restrictive atmosphere rather than the positive outlook that could help patients pursue a better quality of life. The need for constant monitoring and tracking of the body increases adherence pressure and creates a heightened focus on hemoglobin A1c levels. Patients can also initially find diabetes easy to underestimate and ignore, adopting an acute rather than chronic mindset that makes it difficult to take ownership in the long term.

Acute events often are perceived by patients as inevitable consequences of living with chronic conditions. For example, 1 patient with ACS commented: “I learned that even without heart disease, unhealthy habits as far as eating and exercise can lead to stroke and heart failure. This really opened my eyes. At that point my life changed dramatically. I changed the way I shopped, I started eating healthier meals, and I began exercising regularly.” However, acute conditions may have a longer-term impact on social behaviors. ACS can lower social and career performance expectations; patients may feel forced out of the “rat race” by their condition or voluntarily take steps to reduce work-related burdens and pressures, as indicated by 1 patient who commented: “I see people around me going to work and realize how much my health prohibits me from working.”

Based on these moments and experiences, patients may feel that rehabilitation is a challenge and may not fully comprehend or appreciate the potential benefits. Once the advantages of rehabilitation are clearly communicated, the offering is compelling to patients if logistical barriers are no issue.

Participants also described the gradual acquisition of self-care expertise as a journey, which can support them in working toward better health. They also see this journey as influenced by universal health-related experiences, where a patient undergoes a cognitive or emotional change regarding their health engagement and self-care. Although each patient journey is unique, we identified 3 modes of engagement that determine the engagement and behavioral patterns in self-care (Figure 1).

Underlying these modes are clusters of experiences called “experience domains.” At different points in the patient journey, health-related experiences from any domain can affect the patient’s mode of engagement, in a process of fluid exchange and even overlap between domains.

Self-care can be seen as an ongoing process of gaining expertise on how to deeply incorporate management routines into patients’ day-to-day lives (Figure 2). Over the years of encountering the full breadth of health care–related experiences, patients work out the best ways to manage their conditions in the context of their individual lives. This process takes patients from awareness of the need for self-care, through acquiring the necessary practices and tools to learning how to use them successfully. A set of applicable self-care practices are developed, which become habits in the form of routines. Highly engaged patients eventually become informal self-care experts on their own body and health.

**Figure 1.** Patient modes of engagement and experience domains. HCP: health care professional.

![Diagram of patient modes of engagement and experience domains](https://formative.jmir.org/2024/1/e43683/figure1.png)
Interviews and patient log data indicate that a patient’s willingness to engage in self-care and gain expertise may be shaped by the cultural context in which they reside. Some of the participants in the United States expressed a growing mistrust of medical expertise and institutions, with a lack of medical insurance also disrupting the continuity of care. For example, patients from the United States shared: “Never trust someone just because they have a medical degree.” And “I’m on Medicare and Medicaid, I’m disabled, cannot work and now I’m getting paid back for the things I used to say because the clinic won’t cover it.” Effective self-care may also present challenges in countries such as Germany, where many health care systems rely primarily on paper-based clinical records, owing to historical and ongoing data privacy concerns. This impacts the patient experience. A man with T2D from Germany shared: “I am fortunate that my two doctors are located in one facility and can therefore coordinate closely. All the information [about my treatment] is centrally stored and can be viewed at the facility at any time.” Pseudoscientific or alternative approaches to health care, such as homeopathic or organic products as well as spiritual practice, also remain prevalent in Germany and Taiwan; however, participants in Taiwan expressed a high level of trust in HCPs’ expertise and authority, with 1 participant commenting: “I’ll strictly follow the doctor’s advice as best as possible.”

These differing experiences across countries was a common theme during the survey. Although, patients in the United States expressed a mistrust toward institutions, those surveyed were enthusiastic about trying new digital health solutions. However, this attitude was not reflected in Germany, where privacy concerns limiting the uptake of digital solutions. Many patients remain cautious around such technologies. Further, 1 patient with T2D in Germany stated: “I don’t really feel sick, because I don’t notice anything. I don’t have a different life in terms of ‘before’ or ‘after’ the diagnosis. Actually, I ignore the disease.”

**Theorization of Patient Engagement: the Patient Mind States Model (PMM)**

Based on analyses of the survey through patient interviews, we identified 5 patient mind states according to degree of patient engagement, adherence, and the experience domains that drive engagement (namely, the condition-, context-, and self-driven experiences; Figure 3). These mindsets comprise what we have called the Patient Mind States Model (PMM), which articulates these 5 mind states regarding disease self-care. A mind state is defined as a patient’s mental and emotional attitude toward self-care; these states are not related to age, gender, sociodemographic criteria, or culture of persons living with cardiometabolic conditions. The mind state has a large influence on a patient’s receptivity toward support and their ability to develop more healthy behavior. Patients’ mind states are not constant; shifts in mind state can be driven by external forces such as seasonal cycles and life changes. Self-care engagement and maturity of patients with chronic conditions change with their mindset. The 5 mind states identified are the following:

- **Ignoring:** patients carry on with life as it was prior to their diagnosis; depending on their culture and the extent to which they are stuck in an acute mindset, they may believe that a drug can fix the problem. For example, a 68-year-old woman with T2D from Germany commented: “I don’t really feel sick, because I don’t notice anything. I don’t have a different life in terms of ‘before’ or ‘after’ the diagnosis. Actually, I ignore the disease.”

- **Struggling:** patients feel overwhelmed and anxious; in desperately trying to make sense of what is happening, they can be paralyzed and oscillate between desperation and overambition. A 59-year-old woman with ACS from Germany shared: “I sleep very badly, have fears about my health. I feel completely overwhelmed and don’t know how to get out of this dilemma.”

- **Juggling:** people want to focus on positive aspects of life, which are viewed as part of the healing process; they see themselves as trying to find a balance between their own wants and the demands of their condition. One 64-year-old man with ACS from Germany commented: “Life should still be fun, you have to continue to participate in it, even if you are sick...I do not want to miss on the taste of a beer, a glass of wine or a good meal.”
• Controlling: patients have high familiarity with the effects of food and exercise on their body; they constantly learn more about their condition and have a high use of tools to maintain control and promote a feeling of self-reliance. As noted by a 63-year-old man with T2D from the United States: “Being in touch with my body, it has an equal ‘voice’ in determining what is best for me. This helps me to maintain an aggressive mindset and to immediately determine any abnormalities I might be experiencing and take appropriate action.”

• Reframing: people have learned to control their disease and focus on achieving life goals; they use organic methods and vitamins or supplements combined with physical exercise, breathing techniques and meditation to reduce stress. A 61-year-old man with T2D indicated: “Just because someone has a Medical degree doesn’t mean they don’t make mistakes, listen to their advice, but also do some research for yourself. Never blindly follow anyone or anything but find your own ways.”

These mind states may share common traits through their drivers for engagement, as detailed in Figure 1. Through this survey, each mind state is not exclusively self-, context-, or condition-driven, with each having a predominant motivation.

**Figure 3.** The Patient Mind States Model (PMM)—this model contains patient mind states regarding disease self-care identified in patients with chronic cardiometabolic disease.

<table>
<thead>
<tr>
<th>Ignoring</th>
<th>Struggling</th>
<th>Juggling</th>
<th>Controlling</th>
<th>Reframing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonadherent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adherent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parity adherent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adherent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonadherent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Discussion**

**Principal Findings**

Each mind state in the PMM is associated with specific beliefs and challenges and is susceptible to particular drivers of engagement that can be used to propel effective self-care behavior (Multimedia Appendix 1). In the ignoring mind state, these drivers of engagement include a fear of mortality and sense of urgency, social pressure from loved ones, and relatable role models who defy the stigma. As for the “why me” frame of mind of the struggling mind state, affirmative feedback from HCPs, realistic prognosis that shows possible points for medical intervention, and a sense of urgency that prioritizes the disease serve as drivers of engagement. In the juggling mind state, a mind state that focuses on the positive aspects of life as part of the healing process, clear images of cause and effect, gratification through joyful activities, and companionship with peer patients on their disease journey drive engagement. Moreover, people in the controlling mind state are driven to engage through curiosity for, and excitement about, innovation and what is novel; recognition of progress and being awarded for success; and having a sense of medical expertise. When patients are in the reframing mind state, drivers of engagement include an awareness that certain rules are malleable, deep trust in the individual’s own capabilities, and joy of helping peer sufferers. These emotional and social drivers of engagement across each mind state highlight the need for a holistic approach beyond the current physiological and intellectual drivers of engagement in health care, from gratification through device use to feelings of self-efficacy via immediate feedback and to encourage patient self-care behaviors.

Partitioning a person’s mind state into different stages is not a recent concept, with frameworks such as the transtheoretical model (TTM) examining different behavioral states that evolve over time [36]. The TTM consists of several stages of change and different behavioral processes that drive people to transition to a different state (e.g., contemplation, determination, and action). However, the underlying assumption behind this model that decision-making is linear and unidirectional does not recognize how people may move back and forth between such
states over their lifetime, further deviating from expected behavior change trajectory by exponentially changing social and technological context. The PMM described in this survey allows us to nuance further TTM by its nonlinear nature where patient mind states do not necessarily occur in sequential fashion and may even coexist.

It should also be noted that existing behavioral models are used as tools to support clinical decision-making. In diabetes care, the PAM has been leveraged to predict the potential course of outcomes and how underlying social factors may contribute to activation levels. Mean scores obtained from various PAM instruments, such as PAM-13, may offer a concise summary of a person’s knowledge, skills, and motivation. However, the assessment of drivers underlying activation levels and broader consideration of patient context and its evolution are often decoupled from a PAM assessment. In a recent study on patient activation in individuals with T2D, generic health status topics, distress, and social support were all assessed in questionnaires separate from the primary PAM survey [37]. The PMM may offer complementary perspective for assessment of the context for an individual’s mind state through 1 centralized survey and establish the patient modes of engagement and wider experiences, as detailed in Figure 1. Further studies will be required to elucidate how the PMM could form the basis of a behavioral tool in practice.

Opportunities for Future-Focused Digital Solutions

Digital health care has an important role in chronic patient journey. However, many digital patient support programs may suboptimally tailor and target their support based on these important differences, in part because of a lack of data integration across platforms. Guidance on the day-to-day implementation of digital solutions is lacking, so patients often achieve success through trial and error.

Given a patients’ potential frustration with digital tools, engendering a level of comfort and trust in the technology is an important step to effective self-care. Along with privacy concerns, evidence for digital solutions remains as key challenges in establishing digital health as a viable solution for patient self-care [38-40].

Each patient has a unique experience in self-care, with the previous discussion outlining how we can understand the changing mind state of these individuals and how digital solutions may proffer opportunities to improve self-care.

For chronic care digital interventions, our findings suggest that it is important to tailor support to a patient’s mind state, with personal drivers of engagement potentially leading to optimized patient outcomes, adherence, and self-care expertise. Guiding patients throughout their individual health journey to a life worth living is critical and should be based on the individual, attainable life goals, and intelligently balanced compromises that undergo constant revision in the ever-changing context.

These dynamic patient experience mapping refined with help of the PMM may form a more optimal basis in which to effectively integrate digital solutions that enable and support disease self-care, while considering more holistically the context of those living with cardiometabolic conditions. These insights also may warrant further studies in the field of patient adherence and sustainable behavioral change. It will be of interest to further investigate the underlying motivations behind a change in patient mind state and how digital health care may help move individuals from “struggling” to “controlling” mind states, for instance, and effectively ignite intrinsic patient motivation drivers.

These initial data may form a basis of future studies through the validation and refinement of the PMM and the relationship between patient mind states and chronic disease self-care. In particular, future research should clarify patient self-care behaviors and attitudes toward specific digital health care interventions as a critical part of digital intervention design and development processes and verify that the user experience of participants with these 2 conditions in these 3 countries is consistent with patients in other contexts. Equally, it will be critical to understand more deeply the levers of progressive self-care expertise acquisition and use by patients.

The limitations of this survey include its geographic profile (only 3 countries) and the limited sample size of the patient populations (owing to the specific inclusion criteria). Participants who completed the survey may not be representative of the general patient population, as the survey was conducted via the web and patients participated on a voluntary basis. This approach may have introduced a selection bias, such that only the most motivated or educated patients were included. The educational needs of a representative patient population may have therefore been underestimated. Further, the PMM has been generated based on the inputs from participants who were managing ACS and T2D and has not included people living with cardiovascular chronic conditions alone or type 1 diabetes.

This may result in that the PMM may not fully embrace the whole spectrum of cardiometabolic patient profiles. The focus on the social aspect of survey respondents (role of caregivers in building patient’s self-care motivation) has not been sufficient to frame more distinctly in the PMM. However, this survey had reflected the real-life experiences of patients in different clinical and geographic settings.

Conclusions

Any single journey with an acute or chronic condition is consistently shaped by moments, as well as motivational triggers, which may impact a patient’s growing expertise in self-care. Through this behavioral science survey, a heuristically useful framework has been defined on the underlying nature of how patients engage with self-care, which requires further testing and adaptations. Patients will gradually gain expertise in self-care before acquiring more confidence in proactively using a range of practices and tools. However, this path may be shaped by the patient’s mind state, which will impact treatment adherence and their willingness to engage in self-care practices.

For digital health care solutions to be fully integrated into the patient care journey, it is important to understand how such tools should be tailored to a patient’s mind state and how these states may shift when digital solutions are adopted. It will also be important to understand that such solutions may adapt according to changes in patient mind states.
Acknowledgments

Medical writing support for this paper was provided by Matthew Gunther, a medical writer for Ashfield Health, a UDG Healthcare company. Writing support was funded by Sanofi.

Data Availability

Qualified researchers may request access to documents related to this survey. Further details on Sanofi's data sharing criteria, eligible studies, and process for requesting access can be found online. This survey was funded by Sanofi.

Conflicts of Interest

SDG received consultancy fees from Novartis and Sanofi. CD, CM, JL, and MM are Sanofi employees and may hold shares or stock options in the company. The other authors have nothing to disclose.

Multimedia Appendix 1

Key features of different patient mind states.

References


32. Liska et al. JMIR Form Res 2024 | vol. 8 | e43683 | p.147 https://formative.jmir.org/2024/1/e43683 (page number not for citation purposes)
Abbreviations

ACS: acute coronary syndrome
HCP: health care professional
PAM: Patient Activation Measure
PMM: Patient Mind States Model
T2D: type 2 diabetes
TTM: transtheoretical model
Linguistic Variables and Gender Differences Within a Messenger-Based Psychosocial Chat Counseling Service for Children and Adolescents: Cross-Sectional Study

Zeki Efe, MSc; Sabrina Baldofski, PhD; Elisabeth Kohls, PhD; Melanie Eckert, MA; Shadi Saee, MSc; Julia Thomas, MA; Richard Wundrack, MA; Christine Rummel-Kluge, MD

1Department of Psychiatry and Psychotherapy, Medical Faculty, Leipzig University, Leipzig, Germany
2Department of Psychiatry and Psychotherapy, University Leipzig Medical Center, Leipzig University, Leipzig, Germany
3Krisenchat gGmbH, Berlin, Germany
4Department of Psychology, Chair of Personality Psychology, Humboldt Universität zu Berlin, Berlin, Germany

Corresponding Author:
Christine Rummel-Kluge, MD
Department of Psychiatry and Psychotherapy
Medical Faculty
Leipzig University
Semmelweissstr 10
Haus 13
Leipzig, 04103
Germany
Phone: 49 341 9724464
Email: Christine.Rummel-Kluge@medizin.uni-leipzig.de

Abstract

Background: Text messaging is widely used by young people for communicating and seeking mental health support through chat-based helplines. However, written communication lacks nonverbal cues, and language usage is an important source of information about a person's mental health state and is known to be a marker for psychopathology.

Objective: The aim of the study was to investigate language usage, and its gender differences and associations with the presence of psychiatric symptoms within a chat counseling service for adolescents and young adults.

Methods: For this study, the anonymized chat content of a German messenger–based psychosocial chat counseling service for children and adolescents (“krisenchat”) between May 2020 and July 2021 was analyzed. In total, 661,131 messages from 6962 users were evaluated using Linguistic Inquiry and Word Count, considering the following linguistic variables: first-person singular and plural pronouns, negations, positive and negative emotion words, insight words, and causation words. Descriptive analyses were performed, and gender differences of those variables were evaluated. Finally, a binary logistic regression analysis examined the predictive value of linguistic variables on the presence of psychiatric symptoms.

Results: Across all analyzed chats, first-person singular pronouns were used most frequently (965,542/8,328,309, 11.6%), followed by positive emotion words (408,087/8,328,309, 4.9%), insight words (341,460/8,328,309, 4.1%), negations (316,475/8,328,309, 3.8%), negative emotion words (266,505/8,328,309, 3.2%), causation words (241,520/8,328,309, 2.9%), and first-person plural pronouns (499,698/8,328,309, 0.6%). Female users and users identifying as diverse used significantly more first-person singular pronouns and insight words than male users (both P<.001). Negations were significantly more used by female users than male users or users identifying as diverse (P=.007). Similar findings were noted for negative emotion words (P=.01). The regression model of predicting psychiatric symptoms by linguistic variables was significant and indicated that increased use of first-person singular pronouns (odds ratio [OR] 1.05), negations (OR 1.11), and negative emotion words (OR 1.15) was positively associated with the presence of psychiatric symptoms, whereas increased use of first-person plural pronouns (OR 0.39) and causation words (OR 0.90) was negatively associated with the presence of psychiatric symptoms. Suicidality, self-harm, and depression showed the most significant correlations with linguistic variables.

Conclusions: This study highlights the importance of examining linguistic features in chat counseling contexts. By integrating psycholinguistic findings into counseling practice, counselors may better understand users' psychological processes and provide more targeted support. For instance, certain linguistic features, such as high use of first-person singular pronouns, negations, or
negative emotion words, may indicate the presence of psychiatric symptoms, particularly among female users and users identifying as diverse. Further research is needed to provide an in-depth look into language processes within chat counseling services.

(JMIR Form Res 2024;8:e51795) doi: 10.2196/51795

KEYWORDS

e-mental health; chat counseling; crisis; helpline; linguistic; language; Linguistic Inquiry and Word Count; LIWC; psychiatric symptoms

Introduction

Childhood and adolescence are known for their biological, social, and psychological changes as vulnerable periods, in which young people are at an increased risk for experiencing mental health problems. It is also known that an early age of onset of mental illness is a risk factor for poor mental health conditions in adulthood [1,2]. The use of mental health care services for adolescents and young adults can have a positive influence on their attitudes, beliefs, and behaviors, which are known to be important predictors of their later mental health [3]. A growing number of studies indicate that children, adolescents, and young adults use the internet to seek help for their mental health problems because the digital environment is familiar and easily accessible, offers anonymity, and accommodates their need for independence [4-8]. Nearly all young people aged 12 to 19 years in Germany (94%) own a smartphone [9].

With the increased use of smartphones, text messaging has become the primary communication tool for today’s youth [10]. Studies on text messaging usage with mobile phones have shown that adolescents experience text messaging as a quick, easy, convenient, playful, and inexpensive way of communication [11-14]. In line with this trend, a number of crisis helplines and similar services have begun to offer online support services such as chat or email counseling [15-18]. Studies have shown that adolescents prefer texting to talking when seeking help for mental health problems and find it easier to write than to express serious concerns verbally [18,19]. Recent studies support the acceptance, feasibility, and usability of online support services, especially among young people [7,20].

However, written language lacks nonverbal stimuli. Recent research has shown that facial expressions and prosody have an influence on the recognition of a speaker’s intention in face-to-face communication [21,22]. In fact, in comparison to face-to-face interactions, people report higher levels of miscommunication when texting. This might represent a barrier in messenger-based counseling and may make it difficult for young people to understand and interpret the intentions of online counselors [23,24]. This lack of nonverbal cues can also make it more difficult for crisis line counselors to establish and maintain a therapeutic relationship [25,26]. In some studies, counselors reported greater difficulty and a lower perceived ability to establish a therapeutic relationship in the digital environment compared to a face-to-face counseling or therapy setting [25,27]. In addition to nonverbal stimuli, language usage is an important source of information in the therapeutic context, and the way people use words conveys a great deal of information about themselves and their current situation [28].

Language reflects both conscious and unconscious thoughts and feelings [29]. Linguists distinguish 2 aspects in the study of language: the formal and the content features of language. The formal aspect concerns grammar, syntax, reaction speed, speech tempo, etc, whereas content features consider vocabulary and word choice [30]. The investigation of “lexical diversity” thus allows a better insight into the cognitive diversity of people [31,32].

Thus, in the 1970s, it was evidenced for the first time that language use can be a specific marker for psychopathology, especially depression [30]. It was found that individuals with depression use more first-person singular pronouns (ie, “I,” “my,” “me,” and “mine”) in both spoken and written language [33,34], supporting cognitive theories of depression [33], which indicate that depression is associated with an increased self-focus. Recent research has shown that increased use of certain words, for example, sad (eg, “crying,” “grief,” and “sad”) or sleep (eg, “asleep” and “bed”), correlates positively with higher levels of depressive symptoms [35]. Further studies also found gender differences in language usage. For example, it was found that women tend to use more language related to thoughts, emotions, senses, negations, and verbs in the present or past tense than men [36]. Furthermore, women were shown to be more likely to use first-person singular pronouns than men, which is consistent with the higher prevalence of depression in women [36,37]. Besides first-person singular pronouns and negative emotion words, causation words (eg, “because”) were also found to be used more by people having depression [34,38]. In line with this, research confirmed the Seligman theory for learned helplessness, which postulates that individuals at risk for depression attribute the cause of a negative event as being internal, global, and stable, by showing that young adults with negative attributional styles were more likely to develop clinically significant depression than those without such attributional styles [39]. There are also studies indicating negative attributional styles as predictive factors for developing depressive symptoms when experiencing negative life events [40-42]. Due to the trend of digitalization in mental health care, linguistic investigations have been conducted in the digital environment as well. In doing so, a positive association was found between Twitter posts indicating loneliness and mental health problems of the users [43]. In another study, which examined the language usage of users of loneliness forums, it was found that these users tended to use words associated with sadness or a desire for social contact, that is, their overall language leaned toward words with negative valence [44]. Regarding associations with psychopathology, the newest findings indicate that individuals with depressive symptoms used fewer complex syntactic constructions, such as adverbial phrases, perhaps because these require greater cognitive effort
The population of young smartphone users and texters remains a vulnerable and underserved group in crisis counseling, which is why further research on outcomes and the effectiveness of specific communication and counseling strategies is needed [35,36]. To date, there have been no attempts to examine the chat content of crisis counseling services with regard to their linguistic structure. For this purpose, anonymized chat messages from a messenger-based psychosocial chat counseling service, krisenchat (German for “crisis chat”), were used to examine (1) which linguistic indicators and gender differences can be identified within the messages of chat users and (2) how these linguistic indicators are associated with the presence of psychiatric symptoms. Based on the existing literature, it was hypothesized that female users would be more likely to use first-person singular pronouns, negations, and insight words than male users and users identifying as diverse. Additionally, it was hypothesized that higher use of first-person singular pronouns, negations, negative emotion words, and words indicating causation would be associated with a higher likelihood of the presence of psychiatric symptoms among users.

**Methods**

**Sampling and Data Collection**

For the purpose of this study, anonymized chat data from all users receiving counseling between May 2020 and July 2021 were extracted from the krisenchat database. Data extraction and preparation were performed by authors affiliated with krisenchat (ME, SS, JT, and RW) so that chat content remained within the krisenchat database. The anonymized chat data included metadata on the chat (total number of messages and words sent by users during the whole counseling process, and number of sessions) and information about the user that counselors identified and noted during the counseling process (sociodemographic information, such as gender and age, and topics of users’ concerns). krisenchat counselors were volunteers and had a background in psychosocial studies. In addition, they underwent a structured 2-month training in chat-based counseling. Regarding gender, counselors had 3 options (male, female, and diverse) to mark in their documentation. They were encouraged to record the identified gender and not the biological sex of the users. The gender “diverse” included individuals identifying as nonbinary or diverse, or indicating to be unsure about their gender identity. For more information about the study design and the nature of krisenchat, we referred to the initial evaluation study of krisenchat [20]. Linguistic variables were determined using Linguistic Inquiry and Word Count (LIWC; see below for details).

The sample examined in this study was based on the previous evaluation of krisenchat, in which the sample consisted of those who completed a subsequent feedback survey after the counseling session [20]. Thus, out of a total of 11,031 users in the above-mentioned time period, 6962 (63.1%) completed a feedback survey. The chat messages of these 6962 users were analyzed. In total, 661,131 messages (mean 94.96, SD 259.46) from 26,614 chat sessions (mean 3.82, SD 6.24) with a total word count of 8,872,154 (mean 1274.37, SD 2954.57) were analyzed.

**Ethical Considerations**

The Medical Faculty of the University of Leipzig approved this study on August 3, 2021 (372/21-ek). Users were informed about the data protection and privacy policy of krisenchat when they first contacted the counseling service. The chat counseling only began after confirming the policy with “Yes.” Participants in the study confirmed informed consent via an opt-in function before taking part in the feedback survey.

**Measures**

**Linguistic Variables**

LIWC is a software for dictionary-based quantitative text analysis [29]. LIWC performs an automated 1-word analysis based on a lexicon with more than 80 categories (ie, language variables, descriptors, linguistic dimensions, psychological dimensions, concerns, informal language, and punctuation) including a total of 18,711 words. In 2008, the German version of the lexicon was developed, and good equivalence was confirmed for the majority of LIWC categories [47]. The tool has been used in various studies on personality, social, and clinical psychological frameworks and for the analysis of therapeutic essays, everyday communication, or computer-based communication, and it can therefore be considered a reliable software program for quantitative text analysis [47-52].

LIWC counts the number of words within the lexicon over a whole chat and assigns them to categories. The output file includes all categories of the lexicon. All variables, except summary variables, are expressed as percentages of the total word count of a respective chat. Based on previous findings [34-37,44,53,54], the following linguistic variables were considered in this study: first-person singular (eg, “I,” “me,” and “mine”) and first-person plural pronouns (eg, “we,” “us,” and “our”), negations (eg, “no,” “not,” and “never”), positive emotion words (eg, “love,” “nice,” and “sweet”), negative emotion words (eg, “hurt,” “worried,” and “sad”), cognitive process words such as words related to insight (eg, “think” and “know”), and words related to causation (eg, “because” and “effect”).

**Psychiatric Symptoms**

The presence of psychiatric symptoms was assessed during the counseling process and noted by krisenchat counselors. The identification of psychiatric symptoms was derived from the concerns reported by the users. Counselors distinguished between the presence of the following symptoms: depression, anxiety, suicidality, self-harm, addictive behavior, eating disorders, flashbacks, and obsessive-compulsive symptoms. Additionally, symptoms were summed up into a dichotomous variable “psychiatric symptoms” to indicate the presence or absence of psychiatric symptoms (0, “not present”; 1, “present”).

**Statistical Analysis**

Statistical analyses were performed using IBM SPSS Statistics version 27.0 (IBM Corp). A 2-tailed α value of .05 was applied to statistical testing. First, descriptive statistics were performed for sociodemographic variables and linguistic variables of the total sample. Additionally, Kruskal-Wallis H tests (because of nonnormality of the linguistic variables) were used to identify
gender differences in use, that is, metadata for the number of sessions, messages, and words of each user. Then, a 1-way multivariate analysis of variance (1-way MANOVA) was conducted to test for gender differences in language usage controlling for word count. Gender was considered as an independent variable, and all 7 linguistic variables (ie, first-person singular pronouns, first-person plural pronouns, negations, positive emotion words, negative emotion words, insight words, and causation words) were considered as dependent variables. Post-hoc univariate ANOVAs were conducted separately for every linguistic variable. Bonferroni correction was applied to account for multiple testing. Then, binary logistic regression analysis was conducted to examine the predictive effect of linguistic variables (first-person singular and plural pronouns, negations, positive and negative emotion words, insight words, and causation words) on the presence of psychiatric symptoms.

The amount of explained variance as shown by Nagelkerke $R^2$ was interpreted as follows: $R^2 > 0.20$, “acceptable” or small effect size; $R^2 > 0.40$, “good” or average effect size; and $R^2 > 0.50$, “very good” or large effect size [55]. Additionally, Spearman correlation coefficients ($\rho$) were reported between linguistic variables. Finally, with the aim to examine the deeper relationship between linguistic variables and psychiatric symptoms, explorative Spearman correlations ($\rho$) between all 7 linguistic variables and all categories of psychiatric symptoms (suicidality, self-harm, depression, anxiety, eating disorder symptoms, flashbacks, obsessive-compulsive symptoms, and addictive behavior) were computed and interpreted as follows: $\rho=0.10$, small effect size; $\rho=0.30$, moderate effect size; and $\rho=0.50$, large effect size [56].

### Results

#### Sociodemographic Characteristics

The average user was 17 years old (mean 16.55, SD 3.45 years; range 8-25 years), and most users were female (female: 4988/5978, 83.4%; male: 881/5978, 14.7%; diverse: 109/5978, 1.8%). A large number of all users (4841/6962, 69.5%) contacted the counseling service due to psychiatric symptoms. Further concerns identified were psychosocial distress (eg, school-related problems, family-related problems, bullying, etc; 2370/6962, 34.0%) or emotional distress (eg, grief, lovesickness, anger, and loneliness; 2101/6962, 30.2%) [20].

The users participated in an average of 3.82 (SD 6.24) counseling sessions and sent an average of 94.96 (SD 259.46; range 2-11,512) messages with an average of 1274.37 (SD 2954.57) words throughout the counseling process. Additional testing indicated that there were gender differences in the numbers of sessions ($\chi^2=22.849; P<0.001$), messages ($\chi^2=14.863; P<0.001$), and words ($\chi^2=33.036; P<0.001$). The results are presented in Table 1. Subsequent post-hoc tests indicated that female users attended a significantly higher number of sessions ($z=−3.247; P<0.001$; $r=0.04$) and sent significantly more messages ($z=−5.349; P<0.001$; $r=0.07$) than male users, whereas there were no significant differences between female users and users identifying as diverse (number of sessions: $P=0.13$; number of messages: $P=0.18$; number of words: $P=0.22$). Users identifying as diverse also attended a significantly higher number of sessions ($z=−3.441; P<0.001$; $r=0.04$) and sent significantly more messages ($z=−2.972; P<0.001$; $r=0.04$) and words ($z=−3.639; P<0.001; r=0.05$) than male users.

<table>
<thead>
<tr>
<th>Metadata variables</th>
<th>Male$, mean (SD)$</th>
<th>Female$, mean (SD)$</th>
<th>Diverse$, mean (SD)$</th>
<th>$\chi^2 (df)^b$</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Session count</td>
<td>3.23 (4.17)$^c$</td>
<td>4.17 (6.94)$^d$</td>
<td>5.43 (7.56)$^d$</td>
<td>22.849 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Message count</td>
<td>79.05 (145.20)$^c$</td>
<td>103.74 (287.43)$^d$</td>
<td>189.14 (512.32)$^d$</td>
<td>14.863 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Word count</td>
<td>1068.80 (1965.98)$^c$</td>
<td>1392.55 (3279.22)$^d$</td>
<td>2166.18 (4653.08)$^d$</td>
<td>33.036 (2)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

$a$Reduced sample size owing to missing data on gender.

$b$Test statistic for the Kruskal-Wallis $H$ test.

$c$Different letters indicate significant differences between the groups, while same letters indicate no significant differences between the groups.

### Descriptive Statistics of Linguistic Variables

In the total sample, the mean percentage of first-person singular pronouns among all words of a user during the whole counseling process was 11.59% (SD 2.46%), indicating that on average, more than one-tenth of all words written throughout all chat messages was a first-person singular pronoun (“I,” “me,” “my,” or “mine”). The next most used linguistic categories were positive emotion words (mean 4.85%, SD 1.70%) and insight words (mean 4.05%, SD 1.32%). The mean percentage of negations among all words of a user during the whole counseling process was 3.76% (SD 1.54%). Furthermore, negative emotion words were used with a mean percentage among all words of 3.23% (SD 1.27%). Causation words were used with a mean percentage among all words of 2.93% (SD 1.04%). Finally, first-person plural pronouns were least frequently used with a mean percentage among all words of 0.43% (SD 0.63%).

### Gender Differences in Linguistic Variables

Gender-specific descriptive statistics are displayed in Table 2. A 1-way MANOVA showed statistically significant differences in linguistic variables between genders ($F_{14, 11,932}=8.945$; $P<0.001$; partial $\eta^2=0.01$; Wilk $\Lambda=0.979$). Post-hoc univariate ANOVAs were conducted separately for every linguistic variable. Separate ANOVAs and respective Bonferroni-corrected post-hoc tests showed that when controlling for word count, there were statistically significant differences in the use of first-person singular pronouns between genders ($F_{2,5972}=49.780$; $P<0.001$; $\eta^2=0.01$). Further, female users used significantly more insight words ($z=−4.211; P<0.001$; $r=0.07$) and words ($z=−4.211; P<0.001$; $r=0.04$) than male users, whereas there were no significant differences between female users and users identifying as diverse (number of sessions: $P=0.13$; number of messages: $P=0.18$; number of words: $P=0.22$). Users identifying as diverse also attended a significantly higher number of sessions ($z=−3.441; P<0.001$; $r=0.04$) and sent significantly more messages ($z=−2.972; P<0.001$; $r=0.04$) and words ($z=−3.639; P<0.001; r=0.05$) than male users.
P<.001; partial η²=0.02), with female users (mean difference=0.90, 95% CI 0.68-1.12; P<.001) and users identifying as diverse (mean difference=0.98, 95% CI 0.37-1.56; P<.001) using more first-person singular pronouns than male users, whereas there was no significant difference between female users and users identifying as diverse (mean difference=0.08, 95% CI 0.66 to 0.50; P>.99). Furthermore, the use of negations differed significantly between genders (F₂, 5972=4.915; P=.007; partial η²=0.002), with female users using significantly more negations than male users (mean difference=-0.16, 95% CI 0.03-0.30; P=.01), while there were no significant differences in the use of negations between male users and users identifying as diverse (mean difference=0.329, 95% CI 0.05 to 0.71; P=.12) and between female users and users identifying as diverse (mean difference=0.17, 95% CI 0.20 to 0.53; P=.84). Another significant difference was found in the use of negative emotion words between genders (F₂, 5972=4.505; P=.01; partial η²=0.00), with female users using significantly more negative emotion words than male users (mean difference=0.12, 95% CI 0.01-0.23; P=.04; not significant after Bonferroni correction), while no significant differences in the use of negative emotion words were found between female users and users identifying as diverse (mean difference=0.22, 95% CI 0.07 to 0.52; P=.22) and between male users and users identifying as diverse (mean difference=0.11, 95% CI 0.42 to 0.21; P>.99). Finally, the results showed a significant difference in the use of insight words between genders (F₂, 5972=15.215; P<.001; partial η²=0.01), with female users (mean difference=0.26, 95% CI 0.14-0.37; P<.001) and users identifying as diverse (mean difference=0.41, 95% CI −0.15 to 0.47; P=.007) using significantly more insight words than male users, while no significant differences were found between female users and users identifying as diverse (mean difference=0.16, 95% CI −0.42 to 0.21; P=.67). No overall significant differences were found between genders in the use of first-person plural pronouns (F₂, 5972=3.006; P=.05; partial η²=0.00), positive emotion words (F₂, 5972=0.489; P=.61; partial η²=0.00), and causation words (F₂, 5972=2.434; P=.09; partial η²=0.00).

### Table 2. Gender-specific differences in language usage (N=5978).

<table>
<thead>
<tr>
<th>Linguistic variables</th>
<th>Maleᵃ(n=881), mean (SD)</th>
<th>Femaleᵇ(n=4988), mean (SD)</th>
<th>Diverseᶜ(n=109), mean (SD)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-person singular pronouns</td>
<td>10.81 (2.60)ᵇ</td>
<td>11.71 (2.43)ᶜ</td>
<td>11.79 (2.55)ᶜ</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>First-person plural pronouns</td>
<td>0.47 (0.84)</td>
<td>0.42 (0.59)</td>
<td>0.32 (0.43)</td>
<td>.05</td>
</tr>
<tr>
<td>Negations</td>
<td>3.62 (2.24)ᵇ</td>
<td>3.76 (1.39)cᵇᵈ</td>
<td>3.94 (1.49)cᵇᵈ</td>
<td>.007</td>
</tr>
<tr>
<td>Positive emotions</td>
<td>4.87 (1.80)</td>
<td>4.86 (1.67)</td>
<td>4.68 (1.79)</td>
<td>.61</td>
</tr>
<tr>
<td>Negative emotions</td>
<td>3.13 (1.37)ᵇ</td>
<td>3.24 (1.24)cᵇ</td>
<td>3.02 (1.43)cᵇᶜ</td>
<td>.01</td>
</tr>
<tr>
<td>Insight words</td>
<td>3.81 (1.38)ᵇ</td>
<td>4.07 (1.30)cᵇ</td>
<td>4.20 (1.42)cᵇ</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Causation words</td>
<td>2.99 (1.12)</td>
<td>2.92 (1.02)</td>
<td>3.04 (0.92)</td>
<td>.09</td>
</tr>
</tbody>
</table>

ᵃReduced sample size owing to missing data on gender.
ᵇ,ᶜ,ᵈDifferent letters indicate significant differences between the groups, while same letters indicate no significant differences between the groups.

### Predicting Psychiatric Symptoms by Linguistic Variables

The binomial logistic regression model was statistically significant (χ²=25.0; P=.002), resulting in a small amount of explained variance, as shown by Nagelkerke R²=0.124 (Table 3). Of the 10 variables entered into the regression model, all but 3 contributed significantly to the presence of psychiatric symptoms: first-person singular and plural pronouns, negations, negative emotion words, causation words (all P<.001), and female gender (P=.005), while positive emotion words (P=.08), insight words (P=.90), and diverse gender (P=.57) showed no significant effects. Using first-person plural pronouns was associated with a lower likelihood of reporting psychiatric symptoms (odds ratio [OR] 0.39), as did using more causation words (OR 0.90). In contrast, a higher use of first-person singular pronouns was associated with an increased likelihood of reporting psychiatric symptoms (OR 1.18), as did using more negations (OR 1.11) or negative emotion words (OR 1.15). Finally, being female (OR 1.18) or having a higher age (OR 1.04) was also associated with an increased likelihood of the presence of psychiatric symptoms.
Table 3. Prediction of psychiatric symptoms by language usage (N=6962).

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE</th>
<th>Wald</th>
<th>P value</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.04</td>
<td>0.01</td>
<td>22.04</td>
<td>&lt;.001</td>
<td>1.04 (1.02-1.06)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.16</td>
<td>0.06</td>
<td>7.94</td>
<td>.005</td>
<td>1.18 (1.05-1.32)</td>
</tr>
<tr>
<td>Diverse</td>
<td>0.12</td>
<td>0.21</td>
<td>0.33</td>
<td>.57</td>
<td>1.13 (0.75-1.70)</td>
</tr>
<tr>
<td>First-person singular pronouns</td>
<td>0.05</td>
<td>0.01</td>
<td>15.78</td>
<td>&lt;.001</td>
<td>1.05 (1.03-1.08)</td>
</tr>
<tr>
<td>First-person plural pronouns</td>
<td>−0.94</td>
<td>0.05</td>
<td>301.21</td>
<td>&lt;.001</td>
<td>0.39 (0.35-0.44)</td>
</tr>
<tr>
<td>Negations</td>
<td>0.11</td>
<td>0.02</td>
<td>32.14</td>
<td>&lt;.001</td>
<td>1.11 (1.07-1.16)</td>
</tr>
<tr>
<td>Positive emotions</td>
<td>−0.03</td>
<td>0.02</td>
<td>3.09</td>
<td>.08</td>
<td>0.97 (0.95-1.01)</td>
</tr>
<tr>
<td>Negative emotions</td>
<td>0.14</td>
<td>0.02</td>
<td>43.45</td>
<td>&lt;.001</td>
<td>1.15 (1.11-1.21)</td>
</tr>
<tr>
<td>Insight words</td>
<td>−0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>.90</td>
<td>1.00 (0.96-1.04)</td>
</tr>
<tr>
<td>Causation words</td>
<td>−0.11</td>
<td>0.03</td>
<td>13.80</td>
<td>&lt;.001</td>
<td>0.90 (0.85-0.94)</td>
</tr>
<tr>
<td>Constant</td>
<td>−0.87</td>
<td>0.24</td>
<td>13.07</td>
<td>&lt;.001</td>
<td>0.42 (N/A&lt;sup&gt;d&lt;/sup&gt;)</td>
</tr>
</tbody>
</table>

<sup>a</sup> B: regression coefficient.
<sup>b</sup> Degrees of freedom were 1 for all Wald statistics.
<sup>c</sup> OR: odds ratio.
<sup>d</sup> N/A: not applicable.

The results of an additional correlation analysis between all linguistic variables are reported in Table 4. Among others, significant findings included a negative association between first-person singular pronouns and first-person plural pronouns (ρ=−0.24; P<.001). In line with this, first-person singular pronouns were positively correlated with negations (ρ=0.17; P<.001) and negative emotion words (ρ=0.15; P<.001).
Table 4. Spearman correlation coefficients between linguistic variables (N=6962).

<table>
<thead>
<tr>
<th>Variable</th>
<th>First-person singular pronouns</th>
<th>First-person plural pronouns</th>
<th>Negations</th>
<th>Positive emotions</th>
<th>Negative emotions</th>
<th>Insight words</th>
<th>Causation words</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First-person singular pronouns</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ</td>
<td>1</td>
<td>−0.24&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.17&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.04&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.15&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.29&lt;sup&gt;a&lt;/sup&gt;</td>
<td>−0.02</td>
</tr>
<tr>
<td>P value</td>
<td>__&lt;sup&gt;b&lt;/sup&gt;</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.002</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>22</td>
</tr>
<tr>
<td><strong>First-person plural pronouns</strong></td>
<td>−0.24&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>−0.08&lt;sup&gt;a&lt;/sup&gt;</td>
<td>−0.01</td>
<td>−0.10&lt;sup&gt;a&lt;/sup&gt;</td>
<td>−0.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>−0.07&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>__</td>
<td>&lt;.001</td>
<td>.34</td>
<td>&lt;.001</td>
<td>.01</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Negations</strong></td>
<td>0.17&lt;sup&gt;a&lt;/sup&gt;</td>
<td>−0.08&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>−0.02&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.02</td>
<td>0.13&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.04&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>__</td>
<td>.045</td>
<td>.19</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Positive emotions</strong></td>
<td>0.04&lt;sup&gt;a&lt;/sup&gt;</td>
<td>−0.01</td>
<td>−0.02&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>−0.06&lt;sup&gt;a&lt;/sup&gt;</td>
<td>−0.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.04&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value</td>
<td>.002</td>
<td>.34</td>
<td>.045</td>
<td>__</td>
<td>&lt;.001</td>
<td>.023</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Negative emotions</strong></td>
<td>0.15&lt;sup&gt;a&lt;/sup&gt;</td>
<td>−0.10&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.02</td>
<td>−0.06&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>0.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>−0.07&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.19</td>
<td>&lt;.001</td>
<td>__</td>
<td>.023</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Insight words</strong></td>
<td>0.29&lt;sup&gt;a&lt;/sup&gt;</td>
<td>−0.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.13&lt;sup&gt;a&lt;/sup&gt;</td>
<td>−0.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>0.04&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>&lt;.01</td>
<td>&lt;.001</td>
<td>.023</td>
<td>.023</td>
<td>__</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Causation words</strong></td>
<td>−0.02</td>
<td>−0.07&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.04&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.04&lt;sup&gt;a&lt;/sup&gt;</td>
<td>−0.07&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.04&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
</tr>
<tr>
<td>P value</td>
<td>.22</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>__</td>
</tr>
</tbody>
</table>

<sup>a</sup>Statistical significance.
<sup>b</sup>Not applicable.

Associations Between Linguistic Variables and Psychiatric Symptoms

Finally, an exploratory correlation analysis indicated evidence for the relationship between linguistic variables and various psychiatric symptoms, which are displayed in Table 5. Suicidality, self-harm, depression, and anxiety showed the most significant correlations to linguistic variables. In particular, the use of first-person singular pronouns was positively associated with suicidality (ρ=0.11; P<.001) and self-harm (ρ=0.10; P<.001). The use of first-person plural pronouns was negatively associated with suicidality (ρ=−0.10; P<.001) and depression (ρ=−0.14; P<.001). The use of negations was positively associated with suicidality (ρ=0.18; P<.001) and self-harm (ρ=0.12; P<.001). Finally, the use of negative emotion words was positively associated with depression (ρ=0.15; P<.001) and anxiety (ρ=0.15; P<.001).
Table 5. Spearman correlation coefficients between linguistic variables and psychiatric symptoms (N=6962).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Suicidality</th>
<th>Self-harm</th>
<th>Depression</th>
<th>Anxiety</th>
<th>Eating disorder symptoms</th>
<th>Flashbacks</th>
<th>Obsessive-compulsive symptoms</th>
<th>Addictive behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First-person singular pronouns</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ</td>
<td>0.11&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.10&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.08&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.01</td>
<td>0.05&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.02</td>
<td>0.00</td>
<td>-0.01</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.46</td>
<td>&lt;.001</td>
<td>.08</td>
<td>.75</td>
<td>.33</td>
</tr>
<tr>
<td><strong>First-person plural pronouns</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ</td>
<td>-0.10&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.08&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.14&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.04&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.04&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.01</td>
<td>-0.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.04&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.26</td>
<td>.007</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Negations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ</td>
<td>0.18&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.12&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.06&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.06&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.01</td>
<td>-0.02</td>
<td>-0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.40</td>
<td>.07</td>
<td>.28</td>
</tr>
<tr>
<td><strong>Positive emotions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ</td>
<td>-0.05&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.03&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>.08</td>
<td>.41</td>
<td>.87</td>
<td>.06</td>
<td>.19</td>
<td>.09</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Negative emotions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ</td>
<td>0.04&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.12&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.15&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.02</td>
<td>0.02&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.00</td>
</tr>
<tr>
<td>P value</td>
<td>.002</td>
<td>.03</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.18</td>
<td>.045</td>
<td>.82</td>
</tr>
<tr>
<td><strong>Insight words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ</td>
<td>0.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.00</td>
<td>0.08&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.00&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.04&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value</td>
<td>.005</td>
<td>.73</td>
<td>&lt;.001</td>
<td>.65</td>
<td>.67</td>
<td>.61</td>
<td>.045</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Causation words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ</td>
<td>-0.06&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.03&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.04&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>.007</td>
<td>.03</td>
<td>.10</td>
<td>.13</td>
<td>&lt;.001</td>
<td>.63</td>
<td>.32</td>
</tr>
</tbody>
</table>

<sup>a</sup>Statistical significance.

**Discussion**

**Principal Findings and Comparison With Prior Work**

The findings of this study provide first-time valuable insights into the psycholinguistic characteristics of children, adolescents, and young adults seeking psychosocial support through a messenger-based crisis counseling service (*krisenchat*). Previous findings that examined psycholinguistic characteristics in association with mental health, which however focused on texts (eg, from social media or online therapies), could be identified in the chat context as well. Specifically, linguistic variables were found to be associated with the presence of psychiatric symptoms [43,44,57]. The use of first-person singular pronouns, negations, and negative emotion words increased the likelihood of the presence of psychiatric symptoms, while the use of first-person plural pronouns and causation words was associated with a lower likelihood of the presence of psychiatric symptoms. Female gender was also associated with an increased likelihood of the presence of psychiatric symptoms, which is consistent with the higher prevalence of psychiatric symptoms in women [58]. Gender differences were found, with female users exhibiting more frequent use of certain linguistic features. Previous linguistic analyses, especially in the digital context, such as social media platforms, have been performed with a focus on the presence of psychiatric symptoms or associations with psychiatric symptoms, in particular, depressive symptoms [59]. Linguistic analyses of social media have proven useful in predicting depression, anxiety, loneliness, personality disorders, or other mental health issues [43,57,60-62]. As there are no other comparative studies in this field, the focus of the below comparison of the present results with previous findings relies on correlates of linguistic variables with the presence of psychiatric symptoms, especially depression or anxiety.

**First-Person Pronouns**

Starting with the most frequently used linguistic variable among those examined, an increased use of first-person singular pronouns was associated with a higher likelihood of the presence of psychiatric symptoms. Additionally, it was determined that they were used more often by female users and users identifying as diverse than by male users. In contrast, the linguistic variable of first-person plural pronouns was found to be the least used and did not predict the presence of psychiatric symptoms.
Taking into account the most frequently mentioned concerns among the users of krisenchat (see [20]), involving psychiatric symptoms, such as depression or anxiety, and the higher prevalence of depression in female samples [63], these findings are consistent with previous findings on language usage. Thus, in line with cognitive theories of depression (eg, according to [33]) indicating that depression is associated with an increased self-focus, previous research showed that individuals with depression used more first-person singular pronouns (ie, “I,” “my,” “me,” and “mine”) in both spoken and written language [33,34]. Similarly, first-person plural pronouns (ie, “we,” “us,” and “our”) were used significantly less by depressed individuals, which may be attributed to social isolation or lack of social integration and social engagement in the context of depression [64-66]. However, the increased use of first-person singular pronouns may also be a marker for increased vulnerability to stress and negative emotionality and not directly for depression [67,68].

Negations and Causation Words

Negations were found to be significantly associated with the presence of psychiatric symptoms and were used more frequently by female users than male users or users identifying as diverse. Contradicting the hypothesis, the present results indicate that the use of causation words reduces the likelihood of the presence of psychiatric symptoms, while no significant differences were found between genders. In line with this finding, previous studies indicated that low use of causation words is associated with positive treatment outcomes in treatment for personality disorders [53]. It was found that the use of fewer cognitive words, such as causation words, was associated with a more coherent personal story [69]. A recent study underlined this finding by pointing out that patients having depression tend to use significantly more aligned sentences than bringing them into a logical chain compared with a healthy control group [34]. This cross-sectional study design does not allow to draw conclusions about the trend of the use of causation words throughout the chat counseling. The meaning of these divergent results deserves further specific longitudinal research on the development and change of language use across chat counseling.

Negative and Positive Emotion Words

Elevated use of negative emotion words was associated with an increase in the presence of psychiatric symptoms. Negative emotion words were found to differ between genders, that is, more negative emotion words were used by female users than by male users. This is in line with previous research indicating a significantly higher general use of emotion words by women than men, while men were found to use more anger words [70,71]. This is also in accordance with findings showing that more frequent use of negative emotion words, including anxiety, sadness, and anger words, was positively correlated with higher anxiety and depression levels [54].

Depressive Symptoms

Taken together, the results underline a higher likelihood of depression (or anxiety) in users using more self-focused language (first-person singular pronouns), and more negative and fewer positive (emotion) words. The response style theory of rumination in depression, which could also be proven for worry in generalized anxiety, explains that symptoms of repetitive self-focused negative thinking become habitual over time [72,73]. In terms of language usage, this theory suggests that people with high levels of depression or anxiety might communicate using more self-focused language (ie, more I-related pronouns), and more negative and fewer positive words, and that this tendency may become habitual and outside of conscious awareness. In line with this, linguistic analyses of text-based therapy found reductions in the use of first-person singular pronouns, even though language usage was not being focused on in the treatment [74]. Moreover, changes in the use of positive and negative emotion words and words indicating certainty (eg, “always” and “never”) could be identified during the treatment for depression. Researchers interpret these findings as changes in cognitive processes [75].

Suicidality

In relation to depression, it is also important to keep suicidality in mind because messenger-based chat counseling services are used in acute crises, such as suicidality [76]. In accordance with the present findings, previous studies found that suicidal behavior is associated with the use of more I-related pronouns [77]. Likewise, in accordance with the present findings, previous literary analyses indicated that suicidal poets also used fewer first-person plural pronouns than nonsuicidal poets [66]. In addition, these studies showed that the use of more absolutist language, that is, superlatives and intensifiers (eg, “absolutely,” “completely,” “all,” “none,” etc), was associated with suicidality [60,77,78].

The results of this study indicate that by considering language usage, differences in the user population can be discovered and may also be linked to psychopathology. Thus, language usage should be integrated into the counseling strategy. In the context of computer-based analyses, it was found that in addition to the standardized diagnostic tools used to confirm a psychiatric diagnosis, linguistic research showed that systematic analyses of clients’ language may be used to reliably classify them into diagnostic groups [34]. Additionally, computer-based methods were shown to distinguish persons with depression from other clinical subgroups [79]. Thus, linguistic or, in general, qualitative analyses of text messages seem to be advisable to examine chat-based counseling services in more depth. This is even more true than for social media platforms because text messages are less influenced by social desirability, facilitating more granular visibility into changes in linguistic patterns [80].

Implications for Counselors in Chat-Based Counseling

Multiple implications for psychosocial chat counseling have emerged from the findings. First, counselors may use psycholinguistic analyses as an additional tool for assessing the mental health state of users. By monitoring the usage of specific linguistic features, counselors can identify individuals at risk for psychiatric symptoms and tailor interventions accordingly. For example, in cases of elevated use of first-person singular pronouns, negations, and negative emotion words, particularly for female users and users who identify as diverse, a more in-depth exploration of psychiatric symptoms, especially
depression, anxiety, or suicidality, is recommended. Regardless of the actual cause, it seems advisable to monitor the frequency of first-person singular pronouns throughout the chat history. For example, establishing a word-counting function during chat counseling may provide counselors with additional valuable information. In turn, increased use of first-person plural pronouns as well as causation words can be identified as protective factors, and also expanded and used as such. For example, increased use of first-person plural pronouns on the counselor’s side might be helpful to create a sense of “unity” or belongingness. Taking social belongingness into account as a protective factor for depression, suicidality, or mental disorders in general, it seems advisable to create closeness or sociality through language to initially stabilize the user emotionally, for example, meet their need for attachment or create a base to make further recommendations [81].

Furthermore, it can be helpful to mirror statements of users in the context of active listening (positively paraphrased) or to guide the users through targeted questioning techniques into positively formulated thoughts, goals, behavioral directions, etc (ie, in concrete terms) to avoid negations. This implication is also underscored by the fact that increased use of words referring to expectation, trust, and belongingness was associated with lower depression rates [80,81]. Furthermore, the establishment of logical chains, in the sense of causation, also appears to be of great importance. Thus, attempts should be made to bring concerns, feelings, etc into a logical coherent context in the sense of stimulus-reaction chains. Moreover, the use of insight words shows no influence on the presence of psychiatric symptoms, yet reflective functioning is known to be a protective factor [82]. An emphasis on self-reflection could potentially be integrated into psychoeducation by supporting an understanding of one’s own mind as an aspect of resilience and personal agency that can be fostered through social support and professional help [82]. In practice, repeated questions about what feeling is associated with particular problems or concerns, or identifying thoughts or cognition in terms of one’s own belief patterns, can help to practice such skills.

Strengths and Limitations

To the best of our knowledge, this study is the first to examine language usage in a messenger-based crisis counseling service among youth and young adults. In addition to the strength of the large sample size, this study acts as a reference and comparison for further studies in this area, in part because of the use of the internationally recognized method LIWC. Nevertheless, some limitations have to be taken into account. Owing to the retrospective study design, the data rely on convenience sampling, which limits the generalizability to a more mixed-gender population. It should also be noted that the counseling service is offered in German-speaking countries, which is why cultural and linguistic differences in language usage must be taken into account, and generalizability is limited. An international comparison between counseling services in different countries, cultures, and languages would provide insights into similarities and differences. Assuming that people in crisis reach out to the chat counseling service, high use of emotion words seems somewhat expected. Since the nature of the study was cross-sectional and the words were counted across all chat messages (ie, not in chronological order or within sessions), no data can be provided on the trend of the word count for either positive or negative emotion words, which is why no indications can be derived on whether the word count relates only to the beginning of the chat or also to the progression throughout the chat counseling. Longitudinal studies examining changes in language use across consecutive chat sessions may provide further insights into these associations. Consideration must also be given to the nested data structure, which cannot be clearly read owing to the format used to provide data by LIWC. Therefore, for future studies, in addition to the trend of the word count throughout the chat counseling, the consideration of levels (eg, within a message and during a session) is recommended. Therefore, a qualitative analysis of the chats may be beneficial for providing more in-depth insights into individual language usage as well as concerns. A qualitative approach would also ensure the quality of the classification of the presence of psychiatric symptoms. Likewise, no standardized measurement instruments were used, which in turn opens up further opportunities for future research, for example, implementing symptom-specific questionnaires to examine associations between (changes in) symptom severity and language usage.

Conclusion

This study underlines the options, possibilities, and chances of examining psycholinguistic characteristics in psychosocial online chat counseling services for children and adolescents. The identified associations between specific linguistic features and the presence of psychiatric symptoms provide valuable insights for the development of targeted interventions. By considering psycholinguistic findings in the counseling practice, counselors may enhance their understanding of the psychological processes of users and their interventions to offer a more targeted service for children and adolescents seeking help. Nevertheless, further research is needed to investigate the mechanisms underlying linguistic patterns and explore the effectiveness of linguistic-based interventions. At the same time, this would allow further research on the link between specific indicators and changes in specific psychopathology.

Acknowledgments

We acknowledge support from Leipzig University for open access publishing. The Federal Ministry of Health funded this project (ZMI1-2521FEP001).
Authors’ Contributions

ZE, SB, and CRK designed the study. The data set was prepared by ME, SS, JT, and RW. ZE performed the statistical analysis. ZE and SB drafted the article. All authors approved the final manuscript.

Conflicts of Interest

ZE, SB, EK, ME, SS, JT, and RW confirm no conflicts of interest. ME, SS, JT, and RW are paid employees at krisenchat gGmbH. CRK received lecture honoraria from Recordati and Servier outside and independent of the submitted work.

References


44. Andy A. Studying How Individuals Who Express the Feeling of Loneliness in an Online Loneliness Forum Communicate in a Nonloneliness Forum: Observational Study. JMIR Form Res 2021 Jul 20;5(7):e28738 [FREE Full text] [doi: 10.2196/28738] [Medline: 34283026]


Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Rosie, a Health Education Question-and-Answer Chatbot for New Mothers: Randomized Pilot Study

Quynh C Nguyen¹, PhD; Elizabeth M Aparicio², PhD; Michelle Jasczynski², PhD; Amara Channell Doig², PhD; Xiaohe Yue¹, MSc; Heran Mane¹, BSc; Neha Srikanth³, MSc; Francia Ximena Marin Gutierrez², MSW; Nataly Delcid¹; Xin He¹, PhD; Jordan Boyd-Graber³, PhD

¹Department of Epidemiology and Biostatistics, University of Maryland School of Public Health, College Park, MD, United States
²Department of Behavioral and Community Health, University of Maryland School of Public Health, College Park, MD, United States
³Department of Computer Science, University of Maryland Institute for Advanced Computer Studies, University of Maryland, College Park, MD, United States

Corresponding Author:
Quynh C Nguyen, PhD
Department of Epidemiology and Biostatistics
University of Maryland School of Public Health
4254 Stadium Dr.
College Park, MD, 20742
United States
Phone: 1 301 405 6425
Email: qtnguyen@umd.edu

Abstract

Background: Stark disparities exist in maternal and child outcomes and there is a need to provide timely and accurate health information.

Objective: In this pilot study, we assessed the feasibility and acceptability of a health chatbot for new mothers of color.

Methods: Rosie, a question-and-answer chatbot, was developed as a mobile app and is available to answer questions about pregnancy, parenting, and child development. From January 9, 2023, to February 9, 2023, participants were recruited using social media posts and through engagement with community organizations. Inclusion criteria included being aged ≥14 years, being a woman of color, and either being currently pregnant or having given birth within the past 6 months. Participants were randomly assigned to the Rosie treatment group (15/29, 52% received the Rosie app) or control group (14/29, 48% received a children’s book each month) for 3 months. Those assigned to the treatment group could ask Rosie questions and receive an immediate response generated from Rosie’s knowledgebase. Upon detection of a possible health emergency, Rosie sends emergency resources and relevant hotline information. In addition, a study staff member, who is a clinical social worker, reaches out to the participant within 24 hours to follow up. Preintervention and postintervention tests were completed to qualitatively and quantitatively evaluate Rosie and describe changes across key health outcomes, including postpartum depression and the frequency of emergency room visits. These measurements were used to inform the clinical trial’s sample size calculations.

Results: Of 41 individuals who were screened and eligible, 31 (76%) enrolled and 29 (71%) were retained in the study. More than 87% (13/15) of Rosie treatment group members reported using Rosie daily (5/15, 33%) or weekly (8/15, 53%) across the 3-month study period. Most users reported that Rosie was easy to use (14/15, 93%) and provided responses quickly (13/15, 87%). The remaining issues identified included crashing of the app (8/15, 53%), and users were not satisfied with some of Rosie’s answers (12/15, 80%). Mothers in both the Rosie treatment group and control group experienced a decline in depression scores from pretest to posttest periods, but the decline was statistically significant only among treatment group mothers (P=0.008). In addition, a low proportion of treatment group infants had emergency room visits (1/11, 9%) compared with control group members (3/13, 23%). Nonetheless, no between-group differences reached statistical significance at P<0.05.

Conclusions: Rosie was found to be an acceptable, feasible, and appropriate intervention for ethnic and racial minority pregnant women and mothers of infants owing to the chatbot’s ability to provide a personalized, flexible tool to increase the timeliness and accessibility of high-quality health information to individuals during a period of elevated health risks for the mother and child.

Trial Registration: ClinicalTrials.gov NCT06053515; https://clinicaltrials.gov/study/NCT06053515
Introduction

Background

Maternal morbidity and mortality have remained persistent problems in the United States and disproportionately affect women and birthing people from racial and ethnic minoritized backgrounds owing to embedded racism and bias across the medical and public health systems [1-3]. More concerning, >80% of maternal deaths in 2019 were designated as preventable by the Centers for Disease Control and Prevention’s (CDC’s) maternal mortality review committees [4]. In particular, the perinatal period is associated with high risk of depression and anxiety among mothers and is a leading cause of maternal mortality in the United States [1,4,5]. Other leading causes include hemorrhage, cardiovascular and coronary conditions, and substance use disorders [4,5].

Timely and reliable health information may help to reduce the adverse outcomes during pregnancy and in the postpartum period [6]. Children of single parents, with low household income, of a minority group, or whose parents perceived them as being more susceptible are seen more frequently in the emergency department [7-10]. Health education interventions have been shown to reduce emergency department use among infant caregivers [11]. Seeking health information on websites is common among soon-to-be and new parents; however, the quality of information and sources found on the web about pregnancy, birth, parenting, and maternal health were rated by pregnant women and new parents as having varying quality, or the information found was not sufficiently specific to fully answer questions [12,13]. Currently, some popular programs for these vulnerable populations involve resource-intensive home visits, which face challenges in scaling to assist more mothers owing to staff and cost constraints, or nonpersonalized SMS text messages that may not directly address an individual’s questions [14-18].

Recognizing that facilitating maternal and child health equity in the United States will require intervention at all levels of the socioecological model to address the deficits in medical and public health research and practice, our research team selected an innovative approach. This approach consisted of developing a maternal and child health information chatbot that would be iteratively improved through a multiyear, community-engaged research process. The chatbot addresses some limitations of previous strategies by providing personalized health information based on the users’ needs, is readily available at any time, and can include participants nationwide. Rosie offers timely health information from verifiable websites such as children’s hospitals or the CDC to help parents navigate infant care and find clinically correct information to tackle health issues as they arise. In addition, Rosie offers reminders about preventive care visits for infants (eg, well-baby visits) that can also encourage greater continuity of care, which have been shown to reduce emergency department visits for infants [19].

The research team developed the chatbot, named Rosie, to be able to respond to user questions about parenting, pregnancy, and infant development with vetted, trustworthy web-based sources. Question-answering (QA) chatbots such as Rosie, unlike informational sites and frequently asked questions (FAQs) pages, support their users with personalized responses based on the user’s input. They provide users with the unique opportunity to enter their questions in their own words and receive responses to their questions. We built the corpus or index of maternal and child health information using the information derived from expert sources such as federal agencies (eg, the National Institute for Child and Human Development), hospitals (eg, Mayo Clinic), and professional medical organizations (eg, the American Academy of Pediatrics). These sources provided vital information regarding topics such as pregnancy, parenting, infant development, maternal health, and postpartum care.

Chatbots developed to support maternal mental health and parenting have been shown to be a promising intervention needing further evaluation [20-22]. A mixed methods review of literature led by Chua et al [20] suggests that maternal and child health information chatbots have high acceptance among pregnant women and new parents; however, the reviewed papers noted that both development teams and test users expressed preferences for refining the language used in the responses to be more humanlike and for the chatbots to be familiarized with informal, descriptive language to be more adept at generating answers for users who may be describing symptoms or may not know the medical term for the topic of interest. Recognizing these recommendations from the literature, our research team used a multimethod approach to receive substantive, high-quality feedback from participants when launching our pilot evaluation of our maternal and child health information chatbot, Rosie.

Study Objectives

We conducted an experimental pilot study to examine the feasibility, acceptability, and appropriateness of Rosie for our target audience. The pilot study also allowed us to test all the software and equipment, study protocols, and staff coordination to enable remedies before scaling to a full randomized controlled trial. Pilot data were also used to enhance the accuracy of Rosie’s responses by refining the existing models and fine-tuning the mechanisms and heuristics. The focus was on determining access to the resources and the capability of implementing the components and activities of the intervention as planned. Challenges in the provision of any component or the performance of any activity of the intervention were identified and potential solutions were determined. A sample size of 30 was chosen to be within the typical sample size range of a phase 1 clinical trial, according to the National Institutes
of Health definitions, to conduct the study protocols and elicit participant feedback about the Rosie app with sufficient representation from our target group. Preintervention and postintervention tests were completed to qualitatively and quantitatively evaluate Rosie and describe any changes across key health outcomes including postpartum depression and frequency of emergency room visits to inform the full trial’s sample size calculations. Analyses of these data allowed us to present the preliminary findings, which should be interpreted as preliminary evidence, given that the pilot study was not powered to assess treatment-control differences, and this was not the main objective of the pilot study.

Methods

Development and Functionality of Rosie, the Chatbot

Rosie was customized to meet the needs of the target audience through continuous community feedback. Over the course of 3 years, our research team conducted community listening sessions, >20 community demonstrations of Rosie [23], and focus groups with pregnant women and new mothers of color. With this feedback, we customized Rosie to respond to health topics that mothers requested such as feeding tips, sleep advice, and information about rashes and fevers. To the Rosie app, we also added a set of the most popular questions that were asked by mothers as an FAQs page and provided a list of additional resources (e.g., Supplemental Nutrition Assistance Program) benefits and emergency hotlines). Moreover, we added the requested video library that features things such as how to swaddle a baby, change a diaper, or perform cardiopulmonary resuscitation.

To build Rosie’s robust knowledge base, we collected, scraped, and extracted text from 60 sources, including websites of government agencies, hospitals, and professional medical organizations. A corpus of documents about maternal and infant health was built by scraping text from these vetted web domains using Trafilatura, a Python package and a web document processing tool called Scrapy that extracts text from HTML source code [24]. Each web document was then parsed into approximately 73,000 passages by applying a set of heuristics that retain sentence context. These passages were edited as necessary to serve as answers to the mothers’ questions and were used in a question generation model, probably asked questions (PAQs), to produce likely questions from users. The generated questions and their source passages were reviewed by annotators, who either edited both the question and the passage as necessary or discarded the pairs that were unhelpful, inaccurate, or incomprehensible. The answers augmented the existing knowledge base.

In addition, the research team supplemented the knowledge base by manually writing 350 question-and-answer pairs based on feedback from focus groups and community events with pregnant women and new mothers of color, who asked Rosie questions and identified topics of particular interest (e.g., rashes and infant sleep). Only verified sources of health information were used in Rosie’s knowledge base. Sources with sponsored or commercial content were excluded.

Rosie’s underlying QA system uses an unsupervised, dense passage retrieval model. When users ask questions to Rosie, the retrieval model finds relevant content from the knowledge base that best answers the questions. Rosie also provides a source link in her responses, which can direct users to the websites from where the answer was extracted.

To better communicate with the users and understand their needs, we implemented an intent classification model using Rasa, a conversational artificial intelligence software. This classification model is used to categorize users’ text and respond accordingly. For example, it can identify greetings, thank you messages, and requests for information. It is also able to detect potential mental and physical health emergencies and send alerts via Slack, an instant messaging program, to our team members. Upon detection of a possible health emergency, Rosie sends emergency resources and relevant hotline information. In addition, a study staff member, who is a clinical social worker, reaches out to the participant within 24 hours to follow-up.

Rosie was built using Flutter, an open-source user interface software platform by Google, and designed to be compatible with both iPhone and Android devices. The Rosie mobile app has a log-in page with Google authentication, a chat window page that allows users to ask Rosie questions and rate the answers, an FAQs and resources page, and a medical disclaimer page reminding users that Rosie is an informational tool that does not replace professional medical advice and care (Figure 1).
After downloading and logging into the Rosie app, users are asked to enter their estimated due date or their infant’s date of birth. With the users’ permission, Rosie sends push notifications with daily health tips that are generated based on the due date or birth date, thus allowing for personalized advice based on the week-by-week progression of the user’s pregnancy or infant’s health. The past 7 days’ tips are also saved on an app page for users to refer to if needed. The app development team also created an app monitoring system that can detect server-related issues or app interruptions and notify team members via Slack for prompt troubleshooting and resolution. All conversations between users and Rosie were securely stored in Firebase Database.

Recruitment and Enrollment

This was a prospective randomized controlled pilot study involving a mobile app intervention, Rosie the chatbot. To clarify the methods, we followed the CONSORT (Consolidated Standards of Reporting Trials) checklist (Multimedia Appendix 1). From January 9, 2023, to February 9, 2023, participants were enrolled on a rolling basis until the target sample size was met (N=30) for the 3-month randomized pilot study. Participants were recruited using social media posts, including targeted advertisements, and through partnerships with community-based organizations. In addition, our research assistants contacted mothers who had completed an interest form at a previous Rosie community event or focus group.

Interested potential participants completed a screening questionnaire to determine eligibility. Inclusion criteria included being aged ≥14 years, being a woman of color, and either being currently pregnant or having a baby aged ≤6 months. Research assistants contacted each of the 248 potential participants to complete a brief video call to assess eligibility, explain study details, and obtain informed consent for participation. We identified fraudulent interest forms through video calls and review of interest survey meta-data, including IP addresses, to filter out potential participants who were falsely claiming to meet the inclusion criteria or were residing outside the United States.

Eligible participants were randomized into the control group (a monthly children’s book club) or the treatment group (Rosie, the chatbot) using a web-generated table with 15 slots for each study arm, for a total of 30 enrolled participants. After a participant assigned to the Rosie treatment group was unable to fully enroll owing to technical issues, we recruited an additional participant as a replacement. In addition, a mother in the control group experienced a stillbirth during the pilot study and did not complete the postintervention test. Thus, the final analytic sample was 52% (15/29) Rosie treatment group members and 48% (14/29) control group members (Figure 2). Among the 29 participants who were successfully recruited for the pilot study, 3 (10%) were recruited from partner organizations, 6 (21%) were recruited based on our interest forms at past Rosie events, and the remainder (n=20, 69%) were recruited using social media advertisements.
The control group participants were mailed a children’s board book once a month. The control group was modeled after similar programs across the United States that provide free monthly books to children [25]. Books were selected based on feedback from focus groups and community demonstrations with new parents or pregnant women of color, who expressed a desire to have books featuring diverse families.

At the initial enrollment meeting, our research team provided the Rosie treatment group participants instructions about how to install the Rosie app on their smartphones and provided a walkthrough of how to use the app and how to provide feedback about Rosie’s responses to their questions. We emailed each participant a weekly user engagement summary and sent reminder SMS text messages to encourage the use of the app.

**Data Collection, Outcome Measures, and Analysis**

We assigned each participant an identification number to link pretest and posttest data and to track progress through the pilot study using clinical trial management software. Enrolled participants completed a pretest and posttest Qualtrics survey. Pretest surveys included questions about demographics (maternal age, race and ethnicity, education, household size, and health insurance), whether they were pregnant or parenting a young infant, and their due date or their baby’s birth date (as applicable). Pretest surveys also included the Patient Health Questionnaire–9 (PHQ-9) depression scale [26] and assessed the frequency of emergency room visits for infants.

Posttest surveys administered at the 3-month follow-up assessed pregnancy outcomes (birth weight and gestational age), emergency room visits for infants, PHQ-9 depression scale, and group-specific questions. Rosie treatment group members were asked how often they used Rosie and whether they experienced any of the following issues while using the app (eg, “application crashed,” “took too long to get a response,” “was difficult to use,” and “was not satisfied with the answer[s] to my question[s]”). In contrast, the control group members were asked to rate how much they agreed with the following statements on a 5-point Likert Scale ranging from 5 (strongly agree) to 1 (strongly disagree): the books I received were of good quality, the content of the books I received is a good match for my baby’s needs, the books were helpful to me during my pregnancy or parenting my infant, the books were enjoyable to me during my pregnancy or parenting my infant, participating with the book club was easy for me, and I would recommend the book club to other parents.

Both groups were also asked open-ended questions to obtain qualitative feedback. Rosie treatment group members were asked the following open-ended questions: (1) Besides answering your questions, what other features would you like to see on an application like Rosie? (2) Do you have any concerns about using Rosie? If so, please tell us about them; and (3) Do you have any additional feedback to help us build the best Rosie app possible? In an open-ended question, the control group members were asked to provide any additional feedback or ideas about their experience with the book club.

Descriptive statistics of the study sample were calculated based on group assignment. Pretest and 3-month posttest values were examined for postpartum depression and emergency room visits for infants. To examine the statistical significance of between-group differences at baseline, Fisher exact tests were used for categorical variables and Wilcoxon rank sum tests were used for continuous variables. For the within-group pretest and posttest comparisons, paired 2-tailed t tests and McNemar tests were applied. Furthermore, 2-sample 2-tailed t tests were used to compare the between-group differences in the pretest to posttest changes. Qualitative feedback was organized and

---

**Figure 2.** CONSORT (Consolidated Standards of Reporting Trials) diagram of participant enrollment.
presented separately for the Rosie treatment group and book club control group members.

**Ethical Considerations**
The consent form was read aloud by research assistants to each participant and verbal informed consent was obtained. Written consent was obtained through the completion of an web-based form, and a copy of the consent form was sent via email to each enrolled participant. Participants were also encouraged to use the app and were told that if they were continuously enrolled for 3 months and actively engaged with the app by asking questions to the chatbot at least once a week, they would receive a gift card worth US $50. In addition, participants were told that if, at the end of the 3-month pilot study, they were among the top 20% of active users, defined by the number of unique questions sent to Rosie, they would also receive a tablet preloaded with children’s books. Participants were also given a gift card worth US $15, disbursed through a participant incentive distribution platform, Tango, upon completing the pretest and posttest Qualtrics surveys. The study was reviewed and approved (institutional review board study ID: 1556200) by the institutional review board of the University of Maryland, College Park, based on procedures for studies involving human participants.

**Results**

**Overview**
Baseline key demographic characteristics were not statistically significantly different between the Rosie treatment and control groups (Table 1). The mean age of mothers for both groups was 31.7 (SD 4.7) years. Approximately half of the mothers (9/15, 60%) were pregnant, and the other half (6/15, 40%) took care of young infants. Among those with infants, the mean age of infants was 4 months among Rosie treatment group members and 4.6 months among control group members. Most participants (9/15, 60%) were African American or Black, with the remainder being Asian, Hispanic or Latina, or multiracial (Table 1).

<table>
<thead>
<tr>
<th>Table 1. Total enrollment demographicsa.</th>
<th>Rosie (treatment group; n=15)</th>
<th>Book club (control group; n=14)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participant characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (IQR)</td>
<td>31.7 (30-35)</td>
<td>31.6 (26-35)</td>
<td>.83</td>
</tr>
<tr>
<td><strong>Race and ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td>.14</td>
</tr>
<tr>
<td>Asian</td>
<td>1 (7)</td>
<td>4 (29)</td>
<td></td>
</tr>
<tr>
<td>African American or Black</td>
<td>9 (60)</td>
<td>10 (71)</td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>5 (33)</td>
<td>2 (14)</td>
<td></td>
</tr>
<tr>
<td>Multiracial</td>
<td>2 (13)</td>
<td>1 (7)</td>
<td></td>
</tr>
<tr>
<td>Currently pregnant, n (%)</td>
<td>9 (60)</td>
<td>6 (43)</td>
<td>.36</td>
</tr>
<tr>
<td>Parenting infant, n (%)</td>
<td>6 (40)</td>
<td>8 (57)</td>
<td>.19</td>
</tr>
<tr>
<td>Currently pregnant and parenting infant, n (%)</td>
<td>0 (0)</td>
<td>1 (7)</td>
<td>.29</td>
</tr>
<tr>
<td>Infant age, months (Q1-Q3)</td>
<td>3.7 (2-5)</td>
<td>4.8 (4-6)</td>
<td>.15</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
<td></td>
<td>.37</td>
</tr>
<tr>
<td>High school</td>
<td>2 (13)</td>
<td>2 (14)</td>
<td></td>
</tr>
<tr>
<td>Associate degree</td>
<td>1 (7)</td>
<td>2 (14)</td>
<td></td>
</tr>
<tr>
<td>Bachelor degree</td>
<td>3 (20)</td>
<td>5 (36)</td>
<td></td>
</tr>
<tr>
<td>Master degree</td>
<td>3 (20)</td>
<td>4 (29)</td>
<td></td>
</tr>
<tr>
<td>Professional degree</td>
<td>6 (40)</td>
<td>1 (7)</td>
<td></td>
</tr>
<tr>
<td>Average family size, mean (range)</td>
<td>2.6 (2-3)</td>
<td>2.9 (2-3)</td>
<td>.49</td>
</tr>
</tbody>
</table>

*To examine the statistical significance of between-group differences at baseline, Fisher exact tests were used for categorical variables and Wilcoxon rank sum tests were used for continuous variables.

**Acceptability of Rosie**
More than 87% (13/15) of Rosie treatment group members reported using Rosie daily (5/15, 33%) or weekly (8/15, 53%) across the 3-month study period. Most users reported that Rosie was easy to use (14/15, 93%), and that they received a response from Rosie quickly (13/15, 87%). The remaining issues identified included crashing of the app during attempted use (8/15, 53%), and they were not satisfied with some of Rosie’s answers (12/15, 80%; Table 2).
Table 2. Acceptability statistics for Rosie treatment group (n=15).

<table>
<thead>
<tr>
<th>Questions and response options</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>How often did you use Rosie?</strong></td>
<td></td>
</tr>
<tr>
<td>Monthly or less</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Multiple times a day</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Once daily</td>
<td>5 (33)</td>
</tr>
<tr>
<td>Weekly</td>
<td>8 (53)</td>
</tr>
<tr>
<td><strong>Did you experience any of the following issues while using the app?</strong></td>
<td></td>
</tr>
<tr>
<td>Application crashed</td>
<td>8 (53)</td>
</tr>
<tr>
<td>It took too long to get a response</td>
<td>1 (7)</td>
</tr>
<tr>
<td>It was difficult to use</td>
<td>2 (13)</td>
</tr>
<tr>
<td>I was not satisfied with the answer(s) to my question(s)</td>
<td>12 (80)</td>
</tr>
</tbody>
</table>

Health Results: Quantitative

Pilot results suggested better health outcomes for the Rosie treatment group compared with the control group; however, between-group differences did not reach statistical significance. The estimated change in Rosie participants’ PHQ-9 mean depression scores from baseline to posttest period was −3.66 (SD 4.55) among Rosie treatment group participants compared with −2.77 (SD 4.92) among control group members (Table 3). This decline in depression scores between pretest and posttest period was only statistically significant for the Rosie treatment group (P=.008) and not the control group (P=.07). None of the participants from either group reported any emergency room visits for infants at baseline, but this percentage increased to 23% (3/13) for the control group members versus 9% (1/11) for the Rosie treatment group members (Table 3). Notably, 10 (67%) out of 15 mothers who were pregnant at baseline gave birth by the 3-month posttest period, and it could be a possible reason why emergency room visits for infants increased during the posttest period for both groups.

Table 3. Health and health behavior outcomesa.

<table>
<thead>
<tr>
<th></th>
<th>Rosie (treatment group)</th>
<th>Book club (control group)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Maternal depression scaleb</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest period, mean (SD)</td>
<td>5.33 (4.43)</td>
<td>5.31 (3.33)</td>
<td>N/A</td>
</tr>
<tr>
<td>Posttest period, mean (SD)</td>
<td>1.67 (2.64)</td>
<td>2.54 (2.96)</td>
<td>N/A</td>
</tr>
<tr>
<td>Post-pre change, mean (SD)</td>
<td>−3.66 (4.55)</td>
<td>−2.77 (4.92)</td>
<td>.62</td>
</tr>
<tr>
<td>Values, n (%)</td>
<td>15 (100)</td>
<td>13 (100)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Any emergency visit for infants</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest period, n (%)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>N/A</td>
</tr>
<tr>
<td>Posttest period, n (%)</td>
<td>1 (9)</td>
<td>3 (23)</td>
<td>N/A</td>
</tr>
<tr>
<td>Pre-post change</td>
<td>+9.09</td>
<td>+23.08</td>
<td>.60</td>
</tr>
<tr>
<td>Values, n (%)</td>
<td>11</td>
<td>13</td>
<td>N/A</td>
</tr>
</tbody>
</table>

aFor the within-group pretest and posttest comparisons, paired 2-tailed t tests and McNemar tests were applied. Paired 2-tailed t tests comparing pretest and posttest Patient Health Questionnaire depression scores were statistically different for the treatment group (P=.008). No other within-group comparisons were statistically significant at P<.05. Moreover, 2-sample 2-tailed t tests were used to compare the between-group differences in the pretest to posttest changes. P values assess pre- to postperiod changes for treatment versus control groups.
bThe sample size for emergency room visits was smaller because this outcome was assessed among mothers with infants (excludes currently pregnant mothers during the posttest period).
cN/A: not applicable.

Rosie Results: Qualitative Feedback

The Rosie participants provided considerable qualitative feedback about their experiences (Textbox 1). Participants expressed that they liked having a personal library to ask all their pregnancy and parenting questions, but improvements were needed in both the user experience and the content of responses. Participants commented that the quality of Rosie’s responses to pregnancy-related questions seemed to be low in accuracy compared with questions about infant caretaking. Participants noted that they observed improvements in the app’s functionality as the trial progressed and that the addition of an FAQs library and daily tips about baby’s development were helpful.
Textbox 1. Qualitative feedback about Rosie.

Domain and feedback

Strengths
- “I really enjoyed having a personal library to ask all the questions. Especially because after losing my baby, I didn’t receive a ton of targeted ads since my questions were limited to this app.”
- “I liked it when the daily tips were added to the app. Before that, they were only notifications that I couldn’t go back to and often couldn’t fully read. I also liked when the previous day’s answers became available for viewing, so I didn’t have to screenshot or save links.”
- “I like everything else.”

Points system
- “Additionally, the point system didn’t fit with the original study. We were asked to use Rosie a minimum of once a week with a preference for more frequent use. But then the points were awarded for daily use. It was discouraging to not earn points, but with a newborn at home it was a struggle to do anything other than feed him daily.”

Content and technical concerns
- “Perhaps some answers were not very accurate for the age of my baby; they were for younger or older kids.”
- “Answers were not specific enough, lots of glitches with the app.”
- “The answers to the questions were often inaccurate.”
- “It does not answer me well.”
- “Not efficient. All of the answers take you to the same website. Using the app was a waste of time.”
- “My pregnancy-related questions were often answered inaccurately. The baby questions were mildly better, but if the chatbot is intended for both than it needs more training related to pregnancy symptoms and side effects.”
- “I had some trouble with the push notifications as well. At first, they were accurate for the number of weeks along I was in my pregnancy. Then they started to speed up, telling me that I was as many as two weeks ahead of my baby’s gestational age [e.g., it said I was 38 weeks when I was really 36]. I went in and reset it, using the same due date that I started with, and it continued to be incorrect. The app itself didn’t seem to have the same problem, just the notifications.”
- “Due to the high number of inaccurate responses, I was not motivated to continue using the app. I tried to stick with it, but to be honest this chatbot and the accompanying app have a long way to go before they’re ready for implementation.”
- “Finally, last week I had some issues with the app where it was giving me answers in a mix of text and source code formatting. Everything looked like a hyperlink but the links themselves did not work. I received an update notification, updated the app, and the problem persisted. It did fix itself after a couple of days though.”
- “At times when typing a question at the moment of submitting the keyboard would stay open and would not allow the user to hit submit. I had to exit out the app completely and reopen it and it would work again.”
- “Sometimes I had urgent questions, it could’t be use because it was under maintenance often.”
- “Hard to update.”
- “Rosie crashed a few times [like over multiple days when I tried asking a question].”

Suggestions
- “Sometimes Rosie gave me some answers that were not related to what I wanted to know, I understand sometimes terms can apply to two different things but perhaps Rosie can ask Do you mean this (1) or this (2), and then one chooses what is closer to the question one is asking. It happened to me a couple of times but I don’t remember the specific question.”
- “Answers were not specific enough, lots of glitches with the app.”
- “More accuracy with the answers would be great!”
- “Just implement user feedback.”
- “It would be nice to see some statistics [e.g., x% of kids do x by whatever age].”
- “I was pretty unlikely to visit the website Rosie referred me to. I think I would be more likely to view info right on the screen [even a click box with additional text or pics etc].”
- “Random suicide hotline warning was a little bit abrupt and unexpected.”
- “Answers were unrelated sometimes.”
- “Maybe something more personal with the week we are on if its more geared for pregnancy.”
- “I would say maybe [a] different app for moms and another version for expecting moms only due to the fact that some symptom questions were meant for a child rather than me who is pregnant. Overall the app is a great idea and good help/support for all. Thanks for the opportunity.”
When I ask questions, it didn’t answer my answers so I suggest to add more keywords for more accurate answers.”

“Live chat.”

“Maybe a bilingual app?”

“A way to keep a record of my baby’s weight and height.”

“Citations, more images, voice feature.”

“Chance to chat with other moms and create a community of peer to peer questions and answer library. I really wanted to talk to other moms who had experienced loss early in their pregnancy too.”

“It should have a lot of tips for healthy living.”

“Sometimes I would forget what question I already asked because the previous questions would disappear. It could be nice to have an archive of questions.”

“An actual calendar.”

As a result of the feedback, the development team has added more source websites to Rosie’s knowledge bank and has expanded the FAQs section of the app to include topics such as descriptions of the full schedule of well-baby visits and immunizations in the first 2 years of life. As this pilot study was designed to be part of a broad iterative process, negative, neutral, and positive feedback are all integral parts of refining the app’s functionality and expanding its knowledge base.

Book Club Results: Qualitative Feedback

Participants were extremely positive in their feedback about the book club, with participants rating the books as being of high quality and as a helpful tool for parenting (Textbox 2 and Table 4). All participants (14/14, 100%) agreed that they would recommend the book club to other new parents. Participants appreciated the “diversity and bilingual aspects of the books” and that their infants “really enjoyed the Global Babies book and loved to stare at the faces [presented in the books].” A participant offered a recommendation for an additional children’s book by an author whose works focus on social justice leaders in the United States to be offered in the book club.

Textbox 2. Qualitative feedback about the book club.

<table>
<thead>
<tr>
<th>Domain and feedback</th>
<th>Strengths</th>
<th>Suggestions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“I so much love it.”</td>
<td>“The book[s] were good, but I think there are more popular/exciting book options for babies, especially books featuring babies of color. I am thinking of all the books by Jabari Asim for example. The last book was great though and the overall idea for a book club is fantastic. I loved knowing that new books were coming each month.”</td>
</tr>
<tr>
<td></td>
<td>“I loved the diversity and bilingual aspects of the books.”</td>
<td>“Oh, another thing: I got an automatic message from the book club quite frequently with the same message and it seemed redundant.”</td>
</tr>
<tr>
<td></td>
<td>“Baby really enjoyed the global babies book and loved to stare at the faces.”</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Acceptability statistics for the control group (n=13).

<table>
<thead>
<tr>
<th>Item</th>
<th>Score, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The books I received were of good quality</td>
<td>4.54 (0.14)</td>
</tr>
<tr>
<td>The content of the books I received is a good match for my baby’s needs</td>
<td>4.46 (0.18)</td>
</tr>
<tr>
<td>The books were helpful to me during my pregnancy or parenting my infant</td>
<td>4.38 (0.29)</td>
</tr>
<tr>
<td>The books were enjoyable to me during my pregnancy or parenting my infant</td>
<td>4.58 (0.19)</td>
</tr>
<tr>
<td>Participating with the book club was easy for me</td>
<td>4.77 (0.12)</td>
</tr>
<tr>
<td>I would recommend the book club to other parents</td>
<td>4.85 (0.10)</td>
</tr>
</tbody>
</table>
Health Results: Qualitative

A Rosie treatment group participant and a control group participant experienced pregnancy loss (a miscarriage and a stillbirth) during the study. An unexpected finding from one of the participants, who was assigned to the treatment group, was that using Rosie for maternal and child health information helped in shielding them from some emotional distress after their loss, with our participant stating the following:

I really enjoyed having a personal library to ask all the questions. Especially because after losing my baby, I didn’t receive a ton of targeted ads since my questions were limited to this app.

Recognizing that most of the currently available maternal and child health apps track user interactions for advertisers and feature advertisements, Rosie’s development as a no-cost, advertisement-free app may have additional benefits for mothers who value personal data privacy.

Discussion

Principal Findings

The pilot study demonstrates that Rosie is a feasible, acceptable, and appropriate intervention for pregnant women and new mothers of color. Rosie’s software was able to function with a given set of users and was generally able to generate responses to most of the asked questions. Our study found overall reduction in the PHQ-9 depression scale scores from baseline to the 3-month follow-up across both groups. However, the Rosie treatment group experienced a relatively large reduction from baseline to posttest follow-up. The reduction in maternal depression among both groups may correspond to known trajectories in maternal depression, which identified that depressive symptoms peak around birth and decrease as the infant ages [27,28]. Researchers have found variability in the timing and duration of perinatal and antenatal depression, but some studies have found that women who had depression symptoms during the antenatal period were likely to have more intense symptoms during pregnancy than during the postpartum period and that perinatal depression symptoms decreased over time [29,30]. With an expanded time frame of 12 months planned for intervention delivery during the full randomized controlled trial, the research team will be more able to precisely track trends in depressive symptoms and identify what, if any, congruence exists in our participants and the current literature about maternal depression and other mental health symptoms.

Although our small sample size and study design limit our ability to identify the causal pathways between Rosie and changes in depressive symptoms, our findings indicate that an association may exist between the use of the Rosie app and low maternal depression owing to increased parental confidence about their own health and infant caretaking through increased access to accurate health information. Rosie’s ability to provide rapid, accurate response with high-quality sources may also reduce the cognitive burden that pregnant women and new parents described in previous studies that emphasized that sorting through information, making comparisons, and determining the quality of the source of information were significant stressors. In addition, the Rosie app may also reduce maternal depression because it can help provide support to mothers who may not otherwise have access to many health-related supports and resources.

The low rates of emergency room use for infants in the Rosie treatment group compared with the control group aligns with previous study hypotheses that health information provided by Rosie can decrease acute health care use. Nonetheless, this could have occurred through multiple channels including potentially greater use of preventive health care services and Rosie assisting with the identification of relevant health information or clinical guidelines to support infant care.

The qualitative feedback the Rosie participants provided aligns with the conclusions obtained by Chua et al [20] during their review of maternal health chatbots that the first evaluations of these interventions often yield a need for improvement in the language models to understand the variety of ways in which users may ask questions about their pregnancy and child and to provide more precise and accurate responses to these questions.

New Rosie Features in Response to the Pilot Study

User experiences and feedback about the Rosie app has informed the further development of Rosie and continued precision of the QA model. For each of Rosie’s responses, users were able to click “thumbs up” or “thumbs down” to indicate their satisfaction or dissatisfaction with Rosie’s response to their question, and approximately 35% of questions received this additional level of feedback from users. The team analyzed this feedback and enhanced Rosie’s knowledge base by including topics that were not covered in previous iterations of the QA model and further refined the QA model based on the issues identified by the participants and the research team. We analyzed and discussed these interactions weekly with the goal of improving Rosie and initiating improvements in the user experience.

In addition, as a result of user feedback, we have expanded Rosie’s knowledge bank by >10 folds from 75,000 passages to >1.8 million passages extracted from 400,000 documents from verified health sources such as the CDC, National Institutes of Health, Mayo Clinic, and children’s hospitals. Rosie’s previous knowledge bank was restricted to only maternal and infant care questions, but Rosie users had requested information about topics such as managing chronic conditions, food safety and preparation, mental health, and self-care. The expanded corpus now enables mothers to ask any health-related question.

Strengths

This pilot study adds to the existing literature about chatbots broadly and their application in the context of maternal and child health. The team’s findings specific to reduction in maternal depression will help address one of the CDC-identified preventable causes of maternal death. In addition, low emergency room visits for infants suggest potential improvements in infant care and avoidance of some health crises. Our qualitative findings concur with those of previous studies, showing that improved precision in responses is needed [13,20]. Overall, participants found the chatbot as a helpful tool, and this intervention is delivered in a way that is easily
accessible and usable. They also believe that it is an appropriate and acceptable approach for women of color who are pregnant or parenting an infant to receive reliable information. The feedback from our participants is invaluable in the preparation for scaling to a full randomized controlled trial. The use of a multimethod approach that obtained both quantitative and qualitative feedback resulted in a broad understanding of participants’ experiences and needs and addressed some gaps recognized in previous trials of chatbots designed for improving health knowledge.

Limitations
The recruitment of study participants was conducted primarily through web-based advertisements, potentially yielding a sample of participants who are overall more comfortable with using apps and their phones as their primary way of seeking health information than other women who are pregnant or parenting infants. Our sample, overall, was highly educated and the most (26/29, 90%) had health insurance, which may have reduced our ability to detect the experiences and health information needs of mothers without the same level of education or identify the needs of mothers whose lack of insurance may be associated with more variability in the use of emergency rooms. The small sample size of this pilot study resulted in low statistical power. Several health outcomes were found to be different in the posttest period when making between-group comparisons of the Rosie treatment and control groups, but differences did not reach statistical significance. In addition, detection of other between-group differences or predictive relationships between group assignment or demographic variables and outcomes of interest was limited by the small sample size. It was also not feasible to compare pregnant women and those parenting infants within groups at pretest and posttest periods to determine whether there were statistically significant differences owing to the small sample size. However, our approach helped to accomplish our goals for the pilot study and has facilitated a robust planning process for scaling to the full randomized controlled trial.

Conclusions
This pilot study showed that the prototype of the Rosie app is a feasible and usable innovation during pregnancy and postpartum period. This study provides valuable insight into using chatbots to help pregnant women and new mothers of color access reliable information the moment it is requested. Promising pilot results suggest that chatbots may reduce adverse health outcomes among ethnic and racial minoritized mothers; however, additional evaluation is warranted including a planned randomized clinical trial to evaluate the effects of Rosie on maternal and infant outcomes. If successful, chatbots such as Rosie can help address the existing health disparities in maternal and child health that have important intergenerational and downstream health consequences for the nation.

Acknowledgments
The study reported in this paper was supported by research grants from the National Institute on Minority Health and Health Disparities (QCN and EMA; grant number R01MD016037) and by the National Library of Medicine (QCN; grant number R01LM012849). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Authors do not have any conflicts of interests relevant to this study.

Data Availability
The deidentified versions of data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors’ Contributions
Study conceptualization, design, and implementation were completed by QCN and EMA. MJ and FXMG facilitated participant recruitment, enrollment, and data acquisition activities, with supervision from EMA. Development, build-out, and amendments of natural language process models and information corpus were completed by XY, HM, NPS, and JB-G. QCN, MJ, ACD, XY, HM, and XH analyzed, interpreted, and created the tables and diagrams based on the study’s quantitative data. MJ, ACD, and FXMG analyzed and interpreted the qualitative data. QCN, MJ, ACD, XY, HM, FXMG, and XH completed the analysis of current literature about chatbot interventions and implications of previous studies, provided critical analysis of integrated results, drafted and revised the iterations of the manuscript. QCN, EMA, MJ, ACD, XY, HM, NPS, FXMG, ND, XH, and JB-G edited and reviewed the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
CONSORT (Consolidated Standards of Reporting Trials) checklist.
[PDF File (Adobe PDF File), 1262 KB - formative_v8i1e51361_app1.pdf]
References


**Abbreviations**

CDC: Centers for Disease Control and Prevention
FAQ: frequently asked question
PAQ: probably asked question
PHQ-9: Patient Health Questionnaire–9
QA: question-answering
Clinical Needs Assessment of a Machine Learning–Based Asthma Management Tool: User-Centered Design Approach

Lu Zheng1*, MS, PhD; Joshua W Ohde1*, PhD; Shauna M Overgaard1, PhD; Tracey A Brereton1, MS; Kristelle Jose1, MS; Chung-Il Wi2,3, MD; Kevin J Peterson1, PhD; Young J Juhn2,3,4, MD, MPH

1Center for Digital Health, Mayo Clinic, Rochester, MN, United States
2Precision Population Science Lab, Mayo Clinic, Rochester, MN, United States
3Department of Pediatric and Adolescent Medicine, Mayo Clinic, Rochester, MN, United States
4Mayo Clinic Health System Research, Mayo Clinic, Rochester, MN, United States
*these authors contributed equally

Corresponding Author:
Lu Zheng, MS, PhD
Center for Digital Health
Mayo Clinic
200 1st Street South West
Rochester, MN
United States
Phone: 1 480 758 0664
Email: zheng.lu@mayo.edu

Abstract

Background: Personalized asthma management depends on a clinician’s ability to efficiently review patient’s data and make timely clinical decisions. Unfortunately, efficient and effective review of these data is impeded by the varied format, location, and workflow of data acquisition, storage, and processing in the electronic health record. While machine learning (ML) and clinical decision support tools are well-positioned as potential solutions, the translation of such frameworks requires that barriers to implementation be addressed in the formative research stages.

Objective: We aimed to use a structured user-centered design approach (double-diamond design framework) to (1) qualitatively explore clinicians’ experience with the current asthma management system, (2) identify user requirements to improve algorithm explainability and Asthma Guidance and Prediction System prototype, and (3) identify potential barriers to ML-based clinical decision support system use.

Methods: At the “discovery” phase, we first shadowed to understand the practice context. Then, semistructured interviews were conducted digitally with 14 clinicians who encountered pediatric asthma patients at 2 outpatient facilities. Participants were asked about their current difficulties in gathering information for patients with pediatric asthma, their expectations of ideal workflows and tools, and suggestions on user-centered interfaces and features. At the “define” phase, a synthesis analysis was conducted to converge key results from interviewees’ insights into themes, eventually forming critical “how might we” research questions to guide model development and implementation.

Results: We identified user requirements and potential barriers associated with three overarching themes: (1) usability and workflow aspects of the ML system, (2) user expectations and algorithm explainability, and (3) barriers to implementation in context. Even though the responsibilities and workflows vary among different roles, the core asthma-related information and functions they requested were highly cohesive, which allows for a shared information view of the tool. Clinicians hope to perceive the usability of the model with the ability to note patients’ high risks and take proactive actions to manage asthma efficiently and effectively. For optimal ML algorithm explainability, requirements included documentation to support the validity of algorithm development and output logic, and a request for increased transparency to build trust and validate how the algorithm arrived at the decision. Acceptability, adoption, and sustainability of the asthma management tool are implementation outcomes that are reliant on the proper design and training as suggested by participants.

Conclusions: As part of our comprehensive informatics-based process centered on clinical usability, we approach the problem using a theoretical framework grounded in user experience research leveraging semistructured interviews. Our focus on meeting
the needs of the practice with ML technology is emphasized by a user-centered approach to clinician engagement through upstream technology design.

(JMIR Form Res 2024;8:e45391) doi:10.2196/45391

KEYWORDS
asthma; formative research; user-centered design; machine learning (ML); artificial intelligence (AI); qualitative; user needs.

Introduction

Background

Transparency, suitability, and adaptability are cited reasons for the chasm between advances in artificial intelligence (AI) and implementation in health systems [1]. Hindering implementation is a lack of transparency about the data used to make decisions and recommendations [2]. The conceptual suitability of, or aversion to, an algorithm in clinical use is practically governed by a clinician’s autonomous decision to engage with the tool [3]. The adaptability of the algorithm to local patient populations and unique workflows further increases the likelihood of adoption [4]. Logically, a proactive and systematic approach to addressing barriers to transparency, suitability, and adaptability may propel the wider implementation and adoption of AI in patient care [5].

Ultimately, the foundation of this approach is rooted in clinician engagement at the earliest stages of AI development [6]. Determining the user’s complex and diverse requirements for effective machine learning (ML)–based clinical decision support (CDS) tools requires a thorough understanding of the clinical utility of data sources and suitable designs to facilitate contact and response in appropriate settings [7]. This formative usability approach can be achieved through an empathetic and sustained relationship within a multidisciplinary team initiated by early-stage formative research and upstream technology design [8].

In a personalized medical practice aiming to optimize a clinician’s management of asthma, an efficient review of the condition’s characterizing features is critical [9]. Unfortunately, efficient and effective review of these data using electronic health records (EHRs) and timely clinical decisions are impeded by the varied format, location, and workflow of data acquisition, storage, and processing [10]. To support clinicians, we aim to develop an ML-based CDS tool that (1) predicts future risk of asthma exacerbation (AE; risk stratification and resource management), (2) provides this risk evaluation in the context of a summary of relevant information for asthma management (reduction of EHR review burden), and (3) offers options for actionable intervention.

As described in our published work, our AI evaluation plan uses a phased framework (Figure 1 [11]) to address technical performance, usability and workflow, and health impact, and iteratively follows our model documentation steps [12]. This paper describes how we conduct phase 0 and phase 1, highlighting user experience (UX) design and formative research through clinical user shadowing and interviews.

Figure 1. Phased research framework for evaluation of AI applied to the A-GPS project based on Park et al [11]. A-GPS: Asthma Guidance and Prediction System; AI: artificial intelligence; FDA: United States Food and Drug Administration; IRB: institutional review board; NASSS: Nonadoption, Abandonment, Scale-up, Spread, and Sustainability; UX: user experience.
Theoretical Framework

Literature detailing ML-based CDS tool translation indicates an ineffective balance of tool intelligence with explainability, creating gaps in translation, implementation, and accountability [13]. This suggests the need for early engagement with clinical users to mitigate present gaps by gaining a comprehensive and multidisciplinary understanding of present challenges, as well as identifying requirements for development and integration that prioritize both intelligence and practical usability. This formative research required a structured methodology to gather unstructured qualitative data and draw reliable conclusions, especially when diverse aspects and roles are involved. Therefore, UX research methodologies were adopted throughout phase 0 and 1, including the Double Diamond design framework and participatory design method that strategically engaged clinical users to derive unmet needs and identify user requirements [14,15]. The Double Diamond design framework (Figure 2 [16,17]), a graphical guide following the phases of the design process, is widely used to customize and standardize the progression of UX research by incorporating iterative loops and feedback opportunities to progress development [16,18]. This framework supports human-centered design, specifically participatory design, an approach that invites users to be active in the design process as a means of better understanding, meeting, and preempting needs to inform developer efforts [6]. At the “research” stage, the research team conducted interviews and gathered data detailing the challenges and pain points of current pediatric asthma care processes, as well as ideas for innovation. At the “synthesis” stage, the research team categorized data gathered from interviews into themes and reframed findings into opportunities in the form of How-Might-We (HMW) questions [19]. Derived HMW questions served as actionable prompts that acknowledge current challenges requiring solutions and encourage collaborative solution generation representative of relevant clinician’s experiences [19,20]. This UX design process allowed for the communication of user requirements from the perspective of engaged users to provide direct guidance and inform tool development, thereby moving past present challenges of trying to design for users and beginning to design with them at the “ideation” phase [21]. This paper described a detailed walkthrough of the first diamond under the context of Asthma Guidance and Prediction System (A-GPS).

Figure 2. Dan Nessler [16] developed this revamped version of the Double Diamond process based on the British Design Council’s Double Diamond [17] (reproduced from Nessler [16], with permission from Dan Nessler). This paper demonstrated how we went through each step in the first diamond to discover and define “user needs,” thus complete certain aspects on the Phases 0 and 1. HMW: How-Might-We.

Study Objectives

The A-GPS tool is an ML-based CDS tool accessible from “within” the EHR workflow. It aims to summarize all asthma-related context information extracted from the EHR on 1 screen page [9,22]. The tool will be embedded with a functional component of the AE risk model (AE risk model), which applies ML algorithms to predict a patient’s risk of exacerbation in 1 year [23]. This study’s objectives were to qualitatively explore clinicians’ experience with the current asthma management system, identify user requirements to improve algorithm explainability and A-GPS prototype, and identify potential barriers to ML-based CDS system use. Research questions were developed to probe the challenges and pain points of gathering asthma-related information within the current asthma management system, thorough evaluation of clinical team member workflows, user requirements for prototype optimization, algorithm explainability and display.
Methods

Participant Selection and Recruitment
We invited a group of clinicians representing the key roles in asthma management to collect user requirements and listen to their suggestions for future implementation. These roles include physician, nurse practitioner, nurse, and coordinator in primary care and asthma management specialty. Most of them are practicing in the Department of Pediatric and Adolescents Medicine and Family Medicine outpatient practices, where A-GPS will be implemented. A total of 14 participants were recruited by email using a convenience sampling approach and scheduled for a one-on-one, 30 to 60-minute virtual interview.

Data Collection
This formative usability research was directed toward an understanding of user requirements and to facilitate optimal workflow integration, estimate the potential impact of health care delivery factors, and work capacity constraints on achieved benefit. We aimed to collect different facets of qualitative data to identify all stakeholders, understand user needs, probe for optimal tool design to support clinical decision-making and routine workflow for each group in a comprehensive manner. To obtain the clinical context of how the tool will be used in practice, 1 researcher shadowed both sites and described the general patient flow. Next, we scheduled a 60-minute virtual interview with each recruited participant. An introductory statement provided background on the ML-based CDS tool prototype and explained the goal of the interview, developing rapport with interview participants. Each interview session was composed of 2 parts. Part 1 was a routine 30-minute semistructured interview. Interview guides were created for stakeholders, clinicians who were part of the A-GPS project or practice leadership, such as a division or practice chair, and users, defined as those with no stake in A-GPS but are practicing clinicians. Detailed interview guides were attached as supplementary material (Multimedia Appendix 1). Stakeholders were asked specific questions regarding their role as stakeholders in A-GPS and as potential users. Nonstakeholder participants (users) were questioned about their experience and needs as end users. Within part 2 of the interview session, participants were invited to demonstrate an EHR walkthrough on their working computer. During this time, we observed how the clinician routinely uses the system and defines the asthma-related information required to make a medical decision. Additionally, follow-up questions were asked to explore their cognitive process. While part 1 focused on clinicians’ reported problems and individual opinions, part 2 allowed us to observe the current problems and workflow objectively. Each type of data supplemented the other to achieve problem-probing and user needs consolidation.

Data Analysis
UX specialized translational informaticians engaged with practice components to evaluate usability and workflow to determine effectiveness, efficiency, satisfaction, ease of use, explainability, and usage, as described in the AI Evaluation Framework by Overgaard et al [23]. Interviews were transcribed, reviewed, and coded by team members LZ, JWO, KAJ, and TAB. Using a web-based collaboration tool MURAL (Tactivos, Inc) [24], transcripts were coded by identifying emergent themes and categorized into primary research questions asking what, how, where, and when. Other themes included challenges or pain points, barriers to adoption, novel ideas, new insights, and stakeholder considerations. Subthemes were presented as opportunities for change using an HMW question format [19]. Figure 3 provides a brief look into the synthesis and analysis work completed using the MURAL tool. As for the EHR walk-through, we used the data as a reference to make the list of the acquired information in the EHR system.

Figure 3. MURAL (Tactivos, Inc) is an online collaborative tool. Key quotes from participant transcripts were added to the board, using colors to identify participants. Similar or consistent responses across participants were grouped within circles, as shown in the image, to form subthemes. AE: asthma exacerbation; AI: artificial intelligence; ED: emergency department; ER: emergency room; HMW: How-Might-We.
Ethical Considerations
This study was conducted as a quality improvement initiative defined by the Mayo Clinic Institutional Review Board and Mayo Clinic Policy. According to the Code of Federal Regulations, 45 CFR 46.102, the project does not require IRB review. Patient and provider consent were not required, but all participants provided oral consent. To protect participant privacy, the participant’s name and email were used for recruitment purposes only and never linked to audio or transcribed data. Data were used to improve the delivery of health care services at Mayo Clinic.

Results

Participant Characteristics
A total of 14 clinicians were interviewed across 4 Mayo Clinic Health System sites in Minnesota, including Rochester, Red Wing, Albert Lea or Austin, and Kasson. Of those who participated, 7 (50%) were women, 11 (79%) were physicians (Doctor of Medicine or Advanced Practice Registered Nurse) and 3 (21%) were nurses. Their roles included asthma specialist, allergist, pulmonologist, pediatrician, family medicine physician, asthma care coordinator, and medical resident.

Targeted Patient Population and Clinician Users of A-GPS Tool
Clinicians identified opportunities to enhance asthma management for the diagnosed and at-risk population through the use of the A-GPS tool. According to clinicians, the tool would best serve pediatric patients with a diagnosis of asthma and should also aim to cover pediatric patients not officially diagnosed but at high risk of developing asthma, including those identified with the following conditions: symptoms of wheezing or coughing, albuterol or oral steroids use, frequently reported respiratory conditions of pneumonia, lung infections, wheezing, or coughing.

Some patients probably have asthma that we don’t detect, but that’s where I think this tool would be helpful because maybe even though they don’t have a diagnosis of asthma, they’ve had wheezing, or other things listed in their diagnosis and problem list. That would be helpful to avoid missing those people. [P6]

Some kids had been given a bronchodilator because often at 18 months, they present with like viral induced wheezes, and we find it improves with albuterol. So, we get a response to albuterol, and we know that these kids are potentially likely to get asthma, but we typically don’t make that diagnosis until after two. [P4]

Patients with pediatric asthma may be seen by multidisciplinary clinical roles including allergy specialists, pulmonologists, pediatricians, asthma care coordinators, rooming nurses, emergency department physicians, and primary care providers. When identifying proper clinician users of the tool, it was reported that any role that needs to provide asthma management care in practice would benefit from accessing and using A-GPS tool. To capture potential role-based variance in user requirements, the routine workflows and information needs were asked for each participant. It was found that asthma management is coordinated care by dynamic care teams, however, participants demonstrated a preference for a shared view of the tool to gain a shared understanding of patient cases. Even though the responsibilities and workflows vary among different roles, the core asthma-related information and functions they requested were highly cohesive, which allows for a shared information view of the tool.

Usability and Workflow Aspects of the ML System
In general, clinicians welcomed the integration of the AE risk model into the patient’s EHR. In practice, the prediction results are expected to help facilitate preventive actions to support better asthma management. To accomplish this, the AE risk prediction results cannot simply be in the EHR, it also needs to notify clinicians and prompt the care team to follow-up with patients in an expedited fashion.

If risk prediction results are added and approaching the threshold, and you will get a message letting you know that that is happening, that would be the best way to go. Because then you can prevent the next exacerbation, rather than waiting to see the patient the next time they come to the clinic, by then patient might have been through a couple of exacerbations. [P2]

If we are getting this risk score and especially if it were telling them that this is somebody that is at high risk of relapses and recurrences of episodes, then we can make that effort to reach out to those individuals. That should be flowing in my mind. That should be going to our care teams. [P14]

Despite the goal of being proactive, obtaining clinician’s attention to the right patient at the right time in an acceptable format is an issue. Notification methods were suggested by participants and opinions varied based on roles.

When you open the chart, it be helpful to have that notification sent via an in-basket message so that we’re aware and could follow up sooner. There might be cases where we’re aware that they’re high risk, but we can just delete it if we already have that plan for follow-up. [P9]

Best Practice Advisory (BPA) kind of prompts the provider that some action needs to be taken in these areas. And it might be a nice opportunity. Or one of the things we have is emergency action plans. Some similar way that incorporates into an action that needs to be taken or addressed for this patient. [P6]

In-basket message was mentioned by many clinicians as a common type of active alarm. However, it is necessary to balance effective information delivery and alert fatigue as clinicians, especially physicians, receive various alarms and notifications from multiple channels in their daily work.

I would just like the color coding in the records. I do not know if an in-basket message would be effective because we get a lot of them. If it was, I like the message was really clear and can quickly know what it is for. [P13]
Probably not an alarm for high-risk cases. I can imagine people getting annoyed at that, but if it came up in care gaps on the storyboard, like a reminder of something needing to be reviewed, that would be nice. [P4]

I hate to say in-basket messages because that just generates another inbox that the provider doesn’t have time to handle. I think having these folks show up as high-risk followed up by our care coordinators is the right way of handling this. It should be a trigger to get care coordinators or nurses to schedule a visit with the patient, which is more important than notifying the provider. [P14]

Additionally, the alarm or notification should reach clinicians with proper guidance for the next steps.

In BPA, we see alarms as this bright red thing with exclamation points. We’re going to want to act on that, but how do you act on it? Like, does it prompt then if you go into your plan, will it prompt something where you get some choices, like high-risk, you know. Whatever the risk score is, here are some options for you and you click those, and it goes into an order set, and you can order it and you’re done. [P7]

**User Expectations and Algorithm Explainability**

**Expectations and Perceived Impact**

Participants reported that the A-GPS tool is expected to have a positive impact on clinicians’ workflow and patient experiences. Clinicians anticipate a positive impact on usability and workflow by (1) streamlining the review of asthma information, (2) providing patients with “proactive” rather than “reactive” care, (3) empowering patients with a deeper and more personalized understanding of their condition, and (4) improving outcomes. Participants find the tool can be helpful in several situations, such as preparing for an upcoming patient visit, following up on a patient’s condition remotely, and changing or refilling medications based on changes in a patient’s condition. With well-organized asthma information presented at the appropriate time in the workflow, clinicians expect they can save time reviewing information and the care team will have a consistent understanding of patient cases.

Importantly, proactive and preventive care is anticipated with the AE risk model, allowing the prioritization of resources to patients of greatest need, and reduction of AE, emergency department visits, and hospitalizations. Clinicians hope to perceive good usability of the model with the ability to note patients’ high risks and take proactive actions to manage asthma efficiently and effectively. Ideally, with attention-grabbing model output visualization, both patient and caregivers would be more engaged in home-based care after seeing future risks. The potential to further drive higher quality outcomes was identified in the potential to monitor the relationship between patient adherence to medication, symptoms, and other contributing factors.

**Algorithm Explainability**

User requirements of the AE prediction algorithm output emphasized interpretability, logical justification, and validation as is shown in Textbox 1. Specifically, known definitions and levels of risk categories must be explainable, leading to efficient patient classification and resource allocation. Visual indications of severity, such as red, yellow, and green to define high, medium, and low-risk categories paired with a numerical indication were required. Supporting contextual information such as flagging primary features impacting risk prediction and providing a summary of additional asthma management variables were key requirements. Supporting information should be easily accessible and presented as hovering capabilities or links to relevant data (eg, patient history and baseline diagnostics). To assist with algorithm explainability and informing next steps, users required supplemental information on how the prediction score was calculated, bolstered by comparative diagnostics (eg, individual and population baseline values). Clinicians expressed concern regarding accuracy and reliability without significant validation of the model. Requirements included documentation to support the validity of algorithm development and output logic, and a request for increased transparency to build trust and validate how the algorithm arrived at the decision. For successful integration, users require that strategic education and phased implementation must be offered. Education and regular reports on the clear demonstration of value was the preferred strategy to gain an understanding of appropriate A-GPS use and limitations. Examples of stated learning preferences included hands-on training, such as workshops presented at monthly meetings, regular follow-up communication and showcasing of successful use cases, and video tutorials. Importantly, users require a clear demonstration of value to ease adoption, achieved by a phased implementation approach (multisite).
Barriers to Implementation in Context

Accuracy and Reliability

Many participants expect the AI model given by A-GPS will be validated for accuracy and reliability. They also stressed the importance of making the model explainable and transparent to users. Clearly explaining why the model predicted a specific risk score will allow users to understand the logic of the model and its relevance to the patient’s current asthma situation. Without demonstrating validation or providing transparency clinicians will lack trust in the tool and likely not use it, limiting its clinical value.

Clinician and Patient Concerns

Clinicians recognized the potential benefits of A-GPS but voiced several ethical concerns regarding the AE risk model. As is shown in Textbox 2, one concern was the misunderstanding of AI’s role in clinical practice and that AI will override clinician autonomy to make clinical decisions. However, 1 participant asserted that the goal of AI is to provide complementary information and that the clinician would still make the final clinical decisions. A similar concern was the impact the risk prediction model would have on a clinician’s intuition. More specifically, when the AE risk model contradicts the clinician’s professional judgment, the possibility of legal or ethical issues may arise depending on what action the clinician takes.

Textbox 2. Examples of interviewees’ statements about ethical concerns.

- Machine learning introduces a new wrench in things. Because now you’re not giving me a necessarily a recommendation, but you’re giving me insight that might either raise my intuition or lower it. How do you handle having that prediction result legally and ethically and everything else? [P12]
- As a pulmonologist, I am trying to understand how other systematic diseases impact asthma. So, I am also checking tests of other body systems and evaluating by talking to patients. For the populations I am seeing with asthma, hopefully, at some point artificial intelligence could help us, but I just do not see it at this point. [P2]
- This prediction score is not meant to override. This is complimentary information for you. I know you do mental calculations, but this is a data-driven calculation that gives you other complementary information. If there’s a discrepancy, is there anything you are thinking low in emotion, say “hi, just to think about it on this page.” So then, you know, you don't have to go to that page, just look through another page of the sectional summary. [P5]

A patient-specific concern was the potential for unnecessary anxiety and emotional burden on patients and their caregivers when told the AE risk model deems the child at “high risk.” The fear that an asthma event could occur based on a prediction tool that many patients and caregivers may not fully understand may provoke unnecessary changes in the child’s daily activities, as shown in the example below.

Parents may worry about their child if the AI tool says, “high risk of AE” and subsequently change daily decisions, such as not sending their child to school or letting them play outside. [P7]

Discussion

Principal Findings

Principal results are discussed by identified themes. In each theme, we started with “How Might We” questions to inspire discussions on challenges and opportunities.

Usability and Workflow Aspects of the ML System

Challenges and opportunities of usability and workflow aspects of the AI system prompt key questions such as the following: How might we incorporate the A-GPS tool to support workflows of different roles? Clinicians are tasked with a workload that involves increasing patient volumes, more complex diseases, and an overwhelming EHR system. Further, 1 goal of A-GPS is to help alleviate the time clinicians spend in the EHR to find asthma-related information and supplement the clinical decision process required to minimize a patient’s risk of AE. Successfully
incorporating A-GPS into the current workflow of various clinical roles is arguably as important as the tool itself. Participants suggested placing the A-GPS tool in the same location within the EHR and having the same view, regardless of roles. This will allow easier navigation in the EHR within the care team’s current practice as clinicians, nurses, and care coordinators frequently view each other’s screens during patient care.

User Expectations and Algorithm Explainability

Challenges and opportunities of user expectations and algorithm explainability prompt key questions such as the following: (1) How might we communicate A-GPS results in a way that is explainable to patients? Although patients and caregivers were not interviewed, they will receive some level of information from the A-GPS tool communicated to them by the clinician or care team. How the outcome of the AE risk model is explained to patients will be important. Limiting unnecessary anxiety or misunderstandings while still effectively communicating the model results needs to be carefully addressed so families can make appropriate decisions that improve the patient’s outcomes. Properly educating clinicians on how to explain the AE risk model to families is 1 approach that could be tied into the overall education plan for A-GPS. (2) How might we remove barriers to adoption to increase clinician buy-in? The adoption of any new tool or technology rarely goes as planned, it takes time to achieve buy-in from users. To increase clinicians’ buy-in for A-GPS a few barriers should be addressed. First, clinicians need to see that the A-GPS tool is validated, accurate, reliable, and effective at saving time in the EHR. Communicating this data using various educational modalities can increase the reach among clinicians. Second, some clinicians will wait to see the value A-GPS brings to their colleagues before using it themselves. These individuals may be reached by leveraging clinician champions who believe the A-GPS tool improves UX with the EHR and patient outcomes. Ultimately, the best method to facilitate the adoption of the A-GPS tool is to ensure its functionality meets the user’s needs and expectations. Clinicians are more likely to use A-GPS if a clear and concise ML-based CDS tool is created that contains only asthma-related information with easy access to more detailed notes and test results. Moreover, clinicians may use and act on the AE risk model if they trust it, understand what the risk score means for their patient, and understand how the model came to its conclusion for their patient.

Barriers to Implementation in Context

Challenges and opportunities related to implementation and system use in context prompt key questions such as the following: How might we address ethical issues brought on by a difference between the AE risk model and a clinician’s professional judgment? The contradiction between the AE risk model and a clinician’s clinical assessment may pose an ethical and even legal issue. In practice, clinicians may feel pressured to act upon the model’s output in fear of legal challenges even if they believe the patient’s risk of AE is different based on their professional judgment. Although this topic deserves further exploration, it is reasonable to assume in educating clinicians that their clinical judgment supersedes the result from the AE risk model as the model does not take into consideration the multitude of variables the clinician assesses. Moreover, the reason for the development of the AE risk model in this context was to provide supplemental information to improve the care of patients with asthma, not replace the expertise of clinicians. Addressing this upfront with potential users should be a component of A-GPS implementation.

Acceptability, adoption, and sustainability of A-GPS are implementation outcomes that are reliant on the proper design and training as suggested by participants. Without following the guidance of clinicians given in this study the success of A-GPS will be limited, resulting in decreased user satisfaction and clinical effectiveness. To overcome potential barriers to implementation success several priority areas should be met. First, the A-GPS tool needs to be easily accessible within the EHR, in a location that is obvious and consistent across all clinical roles, contains all asthma-related information on a single page, and is visually concise and intuitive. To increase acceptability, the AE risk model needs to be validated for accuracy and reliability and made transparent to users. Transparency is necessary to build trust among clinicians and trust facilitates acceptance and adoption. The AE risk model output needs to be easily interpretable, clearly defined, and intuitive to improve adoption and sustainability. While the risk category output and descriptive statement is essential, an organized display of supporting information popping up in proper format is generally desired as part of the output. Except for the quality of model output, the importance of quality and strategy of education and training cannot be ignored. In a paper by Gordon et al [25], Mayo Clinic took a standardized and efficient approach to provide education and training sessions when implementing a new EHR system. The results demonstrated higher acceptance and confidence among users. This could be a good example for an A-GPS project in terms of successful implementation.

Strengths and Limitations

Conducting interviews with potential users of a new clinical tool not only gains insights into their needs but also encourages buy-in as seeking their input before implementation demonstrates that the research team values user feedback. To our knowledge, this study facilitated buy-in and support among participants as several thanked this study’s team for their efforts to understand user needs. For a broader scope of AI projects, the innovative multi-background collaboration between translational informatics, data science teams and engaged clinical stakeholders guided by the design framework at an early stage provides well-defined user requirements and implementation plans and delivers evidence documentation for upcoming implementation and validation stages. The method can be applied to a wide variety of CDS. In this project, the UX research method was used to engage with clinical stakeholders and prospective end users to gain a comprehensive and multidisciplinary understanding of the role A-GPS is expected to play in pediatric asthma care. Shadowing and interviewing clinical stakeholders were a source of engagement that gathered user requirements from the perspective of potential users, with the objective of informing tool development and translation efforts [21]. Once A-GPS tool is functioning technically, its fit
into the clinical workflow must be evaluated. Moreover, education and documentation must be provided to explain the algorithm and its limitations to effectively translate between the perspectives of experts who created and supported the technology and the perspectives of experts who use the solution to patients. Evaluating the interpretation needs of clinicians, preferences for the display of model output (eg, percentage vs binary threshold), and feature contributions will be assessed based on the data obtained from UX research efforts. Concurrently, the team will also engage clinician stakeholders in the development of model documentation to support explainability [26]. The data obtained from UX research will assist the translational informatics and data science teams in identifying the level of explainability needed to inform and validate the design of A-GPS and supplementally enhance existing workflows [27]. Strategic efforts to promote explainability include applying a documentation framework grounded in scientific research addressing known challenges. This encompasses interdisciplinary best practice reporting requirements that follow phases of model development (prepare, develop, validate, deploy, and maintain) for knowledge continuity throughout the solution’s life cycle [12].

A limitation of this study is the generalizability of design and implementation suggestions as the practice environment of participants has different features. Even though they are all within Mayo Clinic Health Systems, some are working on hospital campuses, while some are from community-based primary care, which led to various user needs. Another limitation is that the perspective of patients and their caregivers were not evaluated in this study. This was purposeful as the intended users of A-GPS are clinicians but the impact the AE risk model may have on families, as stated in the Ethical Considerations section, should be explored further.

**Further Research**

Future research will evaluate the sustainability and scalability of user requirements for enterprise, national, and international adoption of the ML-based CDS tool. Ethical considerations of AI interpretation, patient engagement, and clinician autonomy warrant further investigation. Our research team will conduct multiple studies as we approach the future stages of efficacy and side effects, therapeutic efficacy, and safety and effectiveness planned in our phased comprehensive AI evaluation framework (stage 1). There are more questions to answer in the future: How might we demonstrate to providers the accuracy and reliability of the AI output? How might we define required transparency for AI output? How might we provide an efficient educational module for users and show validation measures to support and explain AI output? How might we handle alarm fatigue, including situations where patients do not respond to providers’ intervention? How might we improve ML-model predicting the risk of AE when the provider’s proper and timely intervention may reduce the performance of the model (eg, positive predictive value)?

**Conclusions**

We aimed to anticipate barriers to the translation of our pediatric asthma management ML-based CDS tool by engaging clinicians in prototype development and optimization leveraging UX research methodologies. In efforts to bolster the transparency, suitability, and adaptability of our solution we qualitatively evaluated user requirements and potential barriers in 3 overarching themes: usability and workflow aspects of the ML system, user expectations and algorithm explainability, and barriers to implementation in context. We presented findings specific to our tool’s risk evaluation in the context of a summary of relevant information for asthma management. This work contributes to phases 0 and 1 of our comprehensive informatics-based AI evaluation frameworks developed by our multidisciplinary team of clinicians, data scientists, translational informaticians, and UX experts at Mayo Clinic [23]. The transparent evaluation and documentation of AI applications in health care enhances clinician and patient trust, supports sharing of AI between hospitals, and increases standards and shared responsibility across the continuum of care. The results of this development study further enhance the model documentation of A-GPS aimed to ensure rigorous evaluation, transparency, and knowledge continuity [12]. A sustainability and scalability evaluation of user requirements will strengthen the potential for national and international adoption of A-GPS.

**Acknowledgments**

We would like to deeply thank Dr Manuel Arteta, Dr Valeria Cristiani, Joy Fladager Muth, Dr Gregory Garrison, Dr Margret Gill, Dr Jason Greenwood, Darcy Hall, Dr Martha Hartz, Dr Julian Lynaugh, Dr Donnchadh O’Sullivan, Dr Sarah Scherger, Dr Eric Schnaith, and Dr David Soma for finding their time for interview and sharing their valuable insights into usability of A-GPS in clinical practice setting. This project was supported by the Center for Digital Health, Precision Population Science Lab, the AI Program of Department of Pediatric and Adolescent Medicine. We acknowledge no generative AI usage in the preparation of this paper.

**Data Availability**

The data sets generated and analyzed during this study are not publicly available due to participants’ privacy, but are available from the corresponding author on reasonable request.

**Conflicts of Interest**

None declared.
References


24. Tactivos Inc. MURAL. URL: https://www.mural.co/ [accessed 2023-12-07]


Abbreviations

A-GPS: Asthma Guidance and Prediction System
AE: asthma exacerbation
AI: artificial intelligence
CDS: clinical decision support
EHR: electronic health record
HMW: How-Might-We
ML: machine learning
UX: user experience

©Lu Zheng, Joshua W Ohde, Shauna M Overgaard, Tracey A Brereton, Kristelle Jose, Chung-il Wi, Kevin J Peterson, Young J Juhn. Originally published in JMIR Formative Research (https://formative.jmir.org), 15.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Exploring Adolescents’ Attitudes Toward Mental Health Apps: Concurrent Mixed Methods Study

Helene Høgsdal1, MSc; Henriette Kyrrestad1, PhD; Marte Rye1, PhD; Sabine Kaiser1, PhD
Regional Centre for Child and Youth Mental Health and Child Welfare - North, Faculty of Health Sciences, Tromsø, Norway

Corresponding Author:
Helene Høgsdal, MSc
Regional Centre for Child and Youth Mental Health and Child Welfare - North
Faculty of Health Sciences
UiT The Arctic University of Norway
Campus Tromsø
Tromsø, 9019
Norway
Phone: 47 77646619
Email: helene.hogsdal@uit.no

Abstract

Background: Adolescence is a critical time in which many psychological disorders develop. Mental health promotion is important, especially during this period. In recent years, an increasing number of mobile apps geared toward mental health promotion and preventing mental illness have been developed specifically for adolescents, with the goal of strengthening their mental health and well-being.

Objective: This study aims to explore adolescents’ attitudes toward mental health apps, as well as the perceived usefulness of mental health apps.

Methods: In this mixed methods study, a total of 183 adolescents (mean age 15.62, SD 3.21 years) answered a cross-sectional questionnaire, with 10 questions (eg, “What do you think about mental health apps in general?”). To complement the quantitative findings, individual interviews were conducted with 9 adolescents, during which they could elaborate on their opinions about mental health apps.

Results: A total of 30% (56/183) of the adolescents in the quantitative study had used a mental health app. Over half of the respondents (77/126, 61.1%) reported that they would use a mental health app if they had a mental health problem as well as that they thought mental health apps were somewhat or very useful (114/183, 62.3%). Availability was the most frequently reported advantage of mental health apps (107/183, 58.8%). Possible associated costs of mental health apps were the most frequently mentioned barrier to their use (87/183, 47.5%). Findings from the interviews also pointed to the importance of the availability of mental health apps as well as their credibility and potential to provide adolescents with autonomy when seeking mental health advice and help.

Conclusions: Overall, the results indicate that adolescents have a positive attitude toward and an interest in mental health apps. However, adolescents are also more or less unaware of such apps, which might be one reason why they are often not used. The findings of this study have important implications for future research on mental health apps and for developers of mental health apps that target young people. The insights gained from this study can inform the development of more effective mental health apps that better meet the needs and preferences of adolescents.

(JMIR Form Res 2024;8:e50222) doi:10.2196/50222

KEYWORDS
mental health applications; mental health; adolescents; adolescent; youth; mobile health; app; apps; application; applications; opinion; opinions; cross sectional; survey; surveys; questionnaire

https://formative.jmir.org/2024/1/e50222
Introduction

Overview

The World Health Organization defines well-being as follows: “Well-being is a positive state experienced by individuals and societies. Similar to health, it is a resource for daily life and is determined by social, economic and environmental conditions” [1]. In today’s society, reference is often made to the importance of working to improve young people’s mental health and well-being as adolescence is a critical, transformative, and challenging time [2]. Research indicates that 10%-20% of adolescents worldwide experience mental health problems [3,4], and in approximately half of all mental disorders, the onset of symptoms occurs before the age of 15 years [5,6]. This illustrates that mental illness in adolescence can have negative effects also on adulthood [6,7]. Today, the need for mental health care and tools for young people is high, and it is argued that provided mental health services for adolescents are not sufficient to keep up with the growing demand [8]. It is therefore important to work toward developing targeted measures and evidence-based interventions that can help young people cope with stress and different challenges [2,9]. Furthermore, it is important to provide supporting tools and resources that can promote healthy habits and behaviors that contribute to overall well-being in adolescence.

In recent years, it has been suggested that mobile devices and the internet represent ideal tools to deliver such interventions to young people [10], and an increasing number of developers and researchers have been following this suggestion [11-14]. In a systematic review, Grist et al [13] highlight the importance of digital interventions developed for young people being of good quality. Furthermore, they point out that the development of digital products, intended for young people, should be designed in collaboration with young people, in order to develop customized and high-quality products for the target group. Thus, a broad understanding of what young people think about specific digital interventions can be beneficial for developing effective products suitable for the target group.

Mental Health Mobile Apps

Mobile apps (ie, apps) are tools where users can receive digital health interventions or information, these health apps have the intention of improving the user’s overall health [12]. Along those same lines, mental health apps are designed to improve mental health and well-being among their users [14-16]. These apps can be designed to address specific mental health disorders [17,18], or they can be more general, focusing on promoting mental health and preventing mental distress through a variety of tools and resources such as emotional self-monitoring and coping strategies [19,20].

One important advantage of mental health apps is their availability [16]. Users can gain access to them anytime and anywhere with a smartphone or other mobile devices. This is particularly useful for individuals who may not have easy access to mental health services due to geographic barriers or for individuals who might benefit from receiving frequent reminders or immediate support [15,16]. Furthermore, mental health apps can reach larger segments of the population compared with in-person therapy, and, in some cases, they have shown promising cost-effectiveness [10,16].

Adolescents’ Engagement With Mental Health Apps

Young people spend a considerable amount of time on their smartphones and use them for entertainment, social purposes, and to access information on different topics, including mental health [21-23]. There are several mental health apps available for adolescents, but as Grist et al [13] point out, the quality of the assessment on these apps is scarce. Research on mental health apps for adolescents indicates that young people are generally satisfied with the access to and ease of use of mental health apps [13,24]. Kenny et al [25] highlight in their study that adolescents prefer mental health apps to be safe, engaging, and easily accessible. The use of mental health apps can also allow people to maintain anonymity when seeking mental health advice or guidance, which is appreciated by young people, who often have a high threshold for help-seeking due to a desire for autonomy and the negative attitudes or stigma surrounding mental health services [26-28].

Nevertheless, the level of user engagement in existing mental health apps is relatively low or moderate and repeated long-time use of mental health apps after downloading them is rare [13,29,30]. This may indicate that despite the fact that young people find mental health apps appealing, it is not sufficient to ensure sufficient use over a long span of time. Regardless of the reasons, it is possible that low user engagement could decrease the overall effectiveness of mental health apps. Research on the effectiveness of mental health apps on adolescent mental health outcomes has revealed mixed results. Some research has failed to illustrate any effect [13,31,32], and others have indicated that mental health apps are promising and have the potential to provide improvement in mental health outcomes [14,33,34]. Yet, there is broad agreement that more research on the effectiveness of mental health apps is needed to fully understand their capacity to improve mental health outcomes [13,32,33]. In order to understand more about the effect mental health apps can have on young people’s mental health, it is also beneficial to examine what young people think about mental health apps and how they should appear in order to be effective for the target group.

This Study

The aim of this study is to explore adolescents’ attitudes toward mental health apps, and their perceived usefulness, using both quantitative and qualitative data. While the quantitative study focuses more on the use and usefulness of these apps as well as on their perceived advantages and disadvantages, the qualitative study examines adolescents’ more general thoughts on mental health apps and how they may be of help to them. Together these data sets provide important insight for the development of mental health apps. The results can help to ensure more user-targeted products and more user engagement, which in turn can contribute to a greater effectiveness of mental health apps.
Methods

Overview
This study used a convergent, parallel QUAN + qual mixed methods design [35]. Quantitative and qualitative data were collected simultaneously, and the qualitative study was used to complement the results of the quantitative study. The quantitative data are based on a cross-sectional survey and the qualitative data are based on semistructured interviews. A total of 2 authors (HK and SK) analyzed the quantitative data, and the other 2 authors (HH and MR) analyzed the qualitative data. The results of the 2 data sets are integrated in the discussion.

Quantitative Study

Participants and Procedure
Adolescents were invited to participate in the study during several events that took place in autumn 2022, including an innovation camp in Mosjøen (a municipality in Northern Norway), “The research days” at UiT The Arctic University of Norway, and the World Mental Health Day event organized by the municipality of Tromsø. To participate, adolescents were asked to scan a QR code, which led to a digital questionnaire [36]. Posters and flyers with the QR code were also distributed at various youth clubs and adolescent health centers in Northern Norway. The questionnaire was anonymous, and participation was voluntary. The final quantitative sample (N=183) consisted of 118 (64.5%) girls, 59 (32.2%) boys, and 6 (3.3%) adolescents that did not specify their gender. Participants were aged between 13 and 19 (mean 15.75, SD 1.65) years; 170 (92.9%) participants were Norwegian and 13 (7.1%) participants reported other nationalities (2.2% were Serbian, 1.1% were Swedish, and 3.8% did not specify their nationality).

Questionnaire

Overview
The questionnaire consisted of 10 questions. Information was collected on demographic characteristics like gender (boy, girl, and other), nationality (Norwegian, Serbian, Swedish, or other), and age (11 to 20 years). There were 5 or 6 questions about mental health apps that are described in the following and an open-ended question where the adolescents could give comments about the topic or the questionnaire in general.

Use of Mental Health Apps
Adolescents were asked 1 or 2 questions developed by Grist et al [37]: “Do you use (or have you used) any apps to help you with mental health problems?” with response options “yes” or “no.” Adolescents who answered “no” were asked a follow-up question: “If you had a mental health problem and there were apps available to help, would you use them?” with response options “yes” or “no.”

Perceived Usefulness of Mental Health Apps
Adolescents were asked: “What do you think of mental health apps in general?” with responses given on a 5-point Likert scale, ranging from (1) “not useful at all” to (5) “very useful.”

Advantages of Mental Health Apps
Adolescents were asked: “What do you think some of the advantages are of using an app for your mental health?” Adolescents were presented with 8 statements developed by Grist et al [37], such as “It is more private” and “I don’t have to talk to someone face to face.” One additional statement, “I can be anonymous” was added to the list by the authors. Adolescents could tick up to 3 statements with which they agreed the most.

Disadvantages of Mental Health Apps
Adolescents were asked: “What do you think some of the barriers are to using an app for your mental health?” Adolescents were presented with 9 statements developed by Grist et al [37], such as “I don’t trust apps,” “I am afraid someone will see the app on my phone,” and “It might cost money.” Adolescents could tick up to 3 statements with which they agreed the most.

Information About Mental Health
To examine where adolescents looked for information about mental health, they were asked: “Which media do you use to find information about mental health?” There were 7 response options as follows: Google, social media (eg, TikTok and Facebook), mental health apps, television, web-based newspapers, podcast or radio, and others, where adolescents could specify a different medium in an open-ended textbox. Adolescents were asked to tick all the media they used with no limitations.

Data Analytical Strategy
Data were analyzed with SPSS (version 29; IBM Corp). Descriptive statistics were calculated and included means, SDs, and frequency distributions. A multiple linear regression analysis was calculated to predict adolescents perceived usefulness of mental health apps, with gender (1=boy, 2=girl), age (11-20 years), and previous use of mental health apps (1=yes, 2=no) as predictors. A significance level of less than .05 was applied.

Qualitative Study

Participants and Procedure
Adolescents who answered the questionnaire were invited to contact the researchers if they wanted to participate in an interview and share their thoughts and opinions about mental health apps. A total of 9 adolescents (all Norwegian) chose to be interviewed (Table 1); 7 elected to be interviewed remotely (via telephone, Microsoft Teams, or Zoom) and 2 elected to be interviewed in person. Before each interview, the adolescents were informed that they were expected to talk about mental health apps, which were defined as mobile apps “developed with the thought of helping people manage their own mental health and wellness” [16]. All interviews were audio recorded on an Apple iPad with the app Diktafon [38], encrypted, and sent directly to Nettskjema [36] before they were transcribed. Two adolescents requested to be interviewed together, hence, 1 interview was conducted with 2 adolescents present (informant 3 and informant 4). These adolescents were considered as individual informants despite being from the same interview.
Table 1. Overview of informants and their experience with mental health apps.

<table>
<thead>
<tr>
<th>Informant</th>
<th>Gender</th>
<th>Age (years)</th>
<th>Experience with mental health apps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informant 1</td>
<td>Girl</td>
<td>15</td>
<td>No</td>
</tr>
<tr>
<td>Informant 2</td>
<td>Boy</td>
<td>16</td>
<td>No</td>
</tr>
<tr>
<td>Informant 3</td>
<td>Girl</td>
<td>15</td>
<td>Yes</td>
</tr>
<tr>
<td>Informant 4</td>
<td>Boy</td>
<td>16</td>
<td>No</td>
</tr>
<tr>
<td>Informant 5</td>
<td>Girl</td>
<td>14</td>
<td>No</td>
</tr>
<tr>
<td>Informant 6</td>
<td>Girl</td>
<td>17</td>
<td>No</td>
</tr>
<tr>
<td>Informant 7</td>
<td>Girl</td>
<td>17</td>
<td>Yes</td>
</tr>
<tr>
<td>Informant 8</td>
<td>Boy</td>
<td>16</td>
<td>Yes</td>
</tr>
<tr>
<td>Informant 9</td>
<td>Girl</td>
<td>14</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Interviews

All interviews were conducted in Norwegian. The interviewer followed a semi-structured interview guide, however, the adolescents were encouraged to talk freely without too much interruption from the interviewer. There were three key questions in the interview guide: (1) “What are your thoughts on mental health applications that are aimed at adolescents?” (2) “What do you think mental health applications can do with regard to adolescents’ knowledge about mental health?” (3) “How do you think mental health applications may be of help to young people?” The interviewer asked follow-up questions when it was natural to go more in-depth on the information the adolescent provided. The interviews varied in duration from 15 to 30 minutes.

Analytic Strategies

The qualitative data were analyzed using thematic analysis inspired by Braun and Clarke [39]. Our analysis was based on a constructionist epistemological position, assuming knowledge is socially constructed and developed through communication and interactions between people [40]. Audio recordings were transcribed by the first author, and transcripts were imported into NVivo [41] to support the organization and coding of the data. Transcripts were read carefully by the first and third authors to identify meaning and patterns in the data. Initial codes were created for the data material, which were then sorted into potential themes. The themes were reviewed and defined in order to identify the essence of each one and the relationship between them. All the codes, themes, and presented quotes are based on data from the Norwegian language. The translation was done using a contextualized hermeneutic approach to translation [42] where the data are presented as close to the original context as possible, hence the quotations in particular are directly translated from Norwegian to English.

Ethical Considerations

The questionnaire in the quantitative study was anonymous and no personally identifiable information was collected. To complete the questionnaire, adolescents had to scan the QR code of their own volition. Moreover, before they could access the questionnaire, the adolescents were presented with information about the study and they were told that they agreed to participate by answering the questions. The adolescents were also informed that they could stop answering the questionnaire at any time without consequences.

The qualitative study was evaluated and approved by the Norwegian Center for Research Data (reference 631424). Adolescents of 16 years or older could consent to take part in the interview themselves, while active parental consent was required from adolescents younger than 16 years of age. Consent was retrieved via a digital consent form in Nettskjema [36]. All interview informants received a cinema gift card with a value of NOK 150 (approximately US $15) as compensation.

Results

Quantitative Study

Of the 183 adolescents who completed the web-based questionnaire, 56 (30.6%) adolescents had used a mental health app. Among adolescents who had not used a mental health app (126/183, 68.9%), approximately half (77/126, 61.6%) said they would use one if they had a mental health problem and there were apps available to help, while 48 (38.4%) adolescents said they would not. When asked “What do you think about mental health apps in general?” 25 (13.7%) adolescents answered that they were not or not very useful, 44 (24.0%) adolescents answered “neither”, and 114 (62.3%) adolescents found them somewhat or very useful (mean 3.6, SD 1.1). The regression analysis to predict the perceived usefulness of mental health apps, identified only gender as a significant predictor ($\beta = .35; P < .001$) indicating that girls perceived mental health apps as more useful than boys.

The most frequently reported advantages of using a mental health app were as follows: “It will always be there when I need it” (107/183, 58.8%), “I don’t have to talk to someone face to face” (83/183, 45.4%), “It is more private” (75/183; 41.0%), and “I can be anonymous” (72/183, 39.3%; Table 2). The most frequently reported barriers were as follows: “It might cost money” (87/183, 47.5%), “I don’t know whether the information in them is accurate or true” (82/183, 44.8%), and “I am afraid someone will see the app on my phone” (78/183, 42.6%; Table 3). The most frequently cited media that adolescents used to find information about mental health were by far “Google” (151/183, 82.5%) and “Social media (for example Facebook, Instagram, TikTok)” (90/183, 49.2%; Table 4).
Table 2. Advantages of using mental health apps (N=183).

<table>
<thead>
<tr>
<th>Statement</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is more private</td>
<td>75 (41.0)</td>
</tr>
<tr>
<td>I don’t have to talk to someone face to face</td>
<td>83 (45.4)</td>
</tr>
<tr>
<td>It will always be there when I need it</td>
<td>107 (58.8)</td>
</tr>
<tr>
<td>I don’t have to wait to get information</td>
<td>45 (24.6)</td>
</tr>
<tr>
<td>I can get support and information whenever I need it</td>
<td>51 (27.9)</td>
</tr>
<tr>
<td>I don’t have to write things like my mood down on paper</td>
<td>13 (7.1)</td>
</tr>
<tr>
<td>It is personal to me</td>
<td>12 (6.6)</td>
</tr>
<tr>
<td>I can be anonymous</td>
<td>72 (39.3)</td>
</tr>
<tr>
<td>Other</td>
<td>6 (3.3)</td>
</tr>
</tbody>
</table>

Table 3. Barriers to using mental health apps (N=183).

<table>
<thead>
<tr>
<th>Statement</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I don’t trust apps</td>
<td>49 (26.8)</td>
</tr>
<tr>
<td>I don’t know whether the information in them is accurate or true</td>
<td>82 (44.8)</td>
</tr>
<tr>
<td>I would prefer to speak to someone face to face</td>
<td>46 (25.1)</td>
</tr>
<tr>
<td>I don’t think apps can help me</td>
<td>36 (19.7)</td>
</tr>
<tr>
<td>I am afraid someone will see the app on my phone</td>
<td>78 (42.6)</td>
</tr>
<tr>
<td>It might cost money</td>
<td>87 (47.5)</td>
</tr>
<tr>
<td>Other</td>
<td>17 (9.3)</td>
</tr>
</tbody>
</table>

Table 4. Most often used media to find information about mental health (N=183).

<table>
<thead>
<tr>
<th>Media</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>151 (82.5)</td>
</tr>
<tr>
<td>Social media (eg, Facebook, Instagram, TikTok)</td>
<td>90 (49.2)</td>
</tr>
<tr>
<td>Mental health mobile apps</td>
<td>23 (12.6)</td>
</tr>
<tr>
<td>Television</td>
<td>28 (15.3)</td>
</tr>
<tr>
<td>Web-based newspapers</td>
<td>8 (4.4)</td>
</tr>
<tr>
<td>Podcast or radio</td>
<td>26 (14.2)</td>
</tr>
<tr>
<td>Other</td>
<td>12 (6.6)</td>
</tr>
</tbody>
</table>

Qualitative Study

Theme 1: Accessibility—The Significance of Approachable Mental Health Apps

A central theme that arose from the interviews, was the experienced and perceived accessibility of mental health apps. The adolescents highlighted that mental health apps were a readily accessible alternative to use in order to find information about health care as well as to get help. Further, they pointed out that mental health apps could be a more accessible helping tool than person-to-person offers. In addition, the adolescents talked about accessibility in terms of knowing about mental health apps’ existence, how appealing they are, and potential costs related to using them.

Many of the young people pointed out that they had neither heard of nor used mental health apps.

I don’t really know about any health-promoting apps that young people use [informant 6]

Some adolescents reflected on reasons why they had not heard of them. For instance 1 adolescent stated that her lack of knowledge of mental health apps may be attributable to marketing campaigns that do not target young people directly or correctly.

I don’t think, for example, that the makers of such apps have managed to reach out to young people [...]. I don’t know if I’m perhaps in the wrong target group or something, but at least it hasn’t really caught on yet, so to speak [informant 1]

The adolescents also pointed out that the information in the mental health apps must be easy to understand to be appealing to them. The adolescents stated that information found on internet might be challenging to comprehend, due to for example
the use of specialized terminology. Therefore, they highlighted that it is important that the information provided in mental health apps is specific to the topic that they want to learn more about.

I think that can be good, if you have a good app where the information is simply explained and that you understand it quite easily, then I think it can be used a lot [informant 1]

An app should be a place where people can read about exactly what they need and not so much medical stuff [informant 1]

The importance of free mental health apps was frequently mentioned in all interviews; adolescents stressed that apps should be equally accessible to everyone regardless of their economic situation. Further, some adolescents stated that charging a fee would weaken the credibility of such apps and call into question the developers’ intentions.

I’ve been on an app where it was obvious that they were going to make money off it. And then the information and everything seemed a bit wrong. Suddenly, I had to pay for a membership in order to continue to use the app; then they lost me [informant 3]

Theme 2: Trustworthiness—The Significance of Credible Mental Health Apps

Most adolescents stated that it is important that app developers are credible and that the information that is provided in apps comes from reliable sources.

It is important that young people find a reliable source. That it’s not just random journalists, or someone who is just trying to get clicks, but that it comes from someone who actually knows something about it. So that you know you are getting the right information [informant 2]

Some adolescents also assumed that it would be easier to judge whether the information on a mental health app is correct, as compared with determining the accuracy of the information on the internet. One adolescent stated that apps developed by established, credible sources would be easier to trust than those created by unknown sources.

I think it would be very easy to just take in everything that’s written there [in a mental health app]. Even the bad stuff. That is why I think that such an app should somehow be under Ung.no [a public information channel for young people in Norway] or something, some reliable sources [informant 5]

Theme 3: Autonomy—Mental Health Apps Can Help Adolescents Help Themselves

Several adolescents stated that self-determination was important to them, so they would appreciate an app, as it would allow them to manage their own mental health and difficulties without interference from adults.

In a way, you want to try to help yourself before it goes so far that you are dependent on others to help you [informant 1]

An app should contain information about what we can do ourselves to get better and where we can ask for help. Sometimes you may just need to get guidelines on how to do it yourself [informant 8]

Further, several adolescents pointed out that an app can be used to ask for help if one does not want to contact in-person services.

I think it’s good that there are apps like that, that you don’t have to go and talk to someone, but that you can have an app on your phone that you have all the time [informant 9]

Most adolescents stated that the apps should include a way to communicate with someone.

If it [an app] can be made so that you can communicate with someone, then perhaps the threshold of asking for help will be even lower [informant 1]

Mm, it would have been good if the app contained something like that, a way to talk to a person online [informant 2]

Adolescents were also interested in interacting within the app in ways that did not involve another person. Some suggested that the app could contain some predefined questions that could be used to give the user a more personalized experience, or that the app could refer them to the right professional.

Several adolescents also pointed out that contacting in-person services can trigger undesirable actions, which might make it easier to trust an app.

And you are probably afraid that it will somehow..., you want to deal with it, but at the same time you are afraid that, for example, a nurse or someone will tell someone else and that there will be a lot of actions at school, and then everyone will know about it [informant 1]

Finally, a frequently mentioned topic in the interviews was the importance of being anonymous in the apps. Adolescents cited the value they placed on anonymity both if they needed information and if they needed actual help with their mental health, as this could make it easier for the adolescents to seek help.

It is important that you can remain anonymous, because not everyone likes to talk about such things if you are recognizable [informant 6]

Discussion

Principal Findings

In this study, we examined adolescents’ attitudes and general thoughts toward mental health apps using a concurrent, mixed methods approach. This research is an important contribution to understanding what young people think about mental health-promoting mobile apps, as well as their thoughts on what they should include and how they should appear.

Overall, the adolescents expressed positive attitudes toward mental health apps, even though few had experience with using them. A large proportion perceived such apps as useful tools.
that can help them cope with normal stresses of life, which is important in order to promote well-being and prevent mental health problems [43,44]. In the quantitative section of the study, availability was the most chosen benefit of using mental health apps. This is not surprising as adolescents in Norway spend a lot of time on their phones and on social media [21]. Privacy, as well as the possibility of remaining anonymous, was the subsequent most chosen advantage. These results are in accordance with those reported in previous literature [16,25,37].

Further, in the interviews, adolescents expressed that mental health apps can be of help at any time of the day, also when in-person services are not available. However, the adolescents expressed, that for mental health apps to be accessible, they need to also be easily understandable and have no costs attached.

It is well-known that adolescents tend to be reluctant to seek help, despite the importance of obtaining help and support for improved well-being [45-47]. In this study, adolescents expressed that mental health apps can lower adolescents’ threshold for asking for help, by always being available and by offering anonymity in the help-seeking process. Several interviewees also highlighted the empowering potential of mental health apps. The adolescents valued how an app could provide them with the ability to help themselves without interference from others. Meeting young people’s need for autonomy is important to ensure that they ask for help when they need it [28]. The findings from this study concur with previous findings, which suggested that adolescents value the autonomy that mental health apps provide [48].

Further, the adolescents emphasized that mental health apps should provide an anonymous way for them to talk about their problems, with either a web-based assistant or a real person. In addition, some adolescents reported fear of someone else seeing a mental health app installed on their phone as a deterrent to using such apps. Anonymity is often mentioned as an important reason why adolescents use web-based services when searching for help or advice regarding mental health [16,49], our findings support that adolescents view anonymity as an advantage of using mental health apps, which agrees with previous findings [16,25].

Another important aspect adolescents highlighted was the difficulty they might have in judging whether the information contained in a mental health app was accurate. Most adolescents reported that they used the internet (ie, Google and social media) to search for mental health-related information. The adolescents stated that mental health apps should be developed by credible organizations or individuals that adolescents already trust. Adolescents also stated that, the information presented in the app should be clear and concise in order to be engaging for them. Previous research have also shown the importance of content and appearance to ensure engaging mental health apps [25,50].

Although adolescents generally have a positive attitude toward the increased development and use of mental health apps, few of the adolescents in our study sample were familiar with such apps. These results are in line with previous research which indicates that adolescents are not highly engaged in mental health apps [13,29]. Indeed, some adolescents believed that they had not heard of such apps due to poor marketing strategies. They stated that, if the apps were meant to target adolescents, then they should be marketed to young people directly. However, there are currently strict rules regarding advertising to children and adolescents in Norway [51], which may pose challenges in devising effective marketing strategies to reach adolescents.

Strengths and Limitations

By including both quantitative and qualitative data, this study provides valuable knowledge and insight into adolescents’ perspectives and opinions on mental health apps. However, there are several limitations that must be taken into consideration. In the quantitative study, the sample size was relatively small and not representative. One should therefore be careful about drawing general conclusions from the findings. However, the results can be an important contribution to the field of research on what adolescents think about mental health apps. Moreover, the interviews were of short duration, which may have influenced the depth of the answers. However, the adolescents were encouraged to speak freely, and they were not interrupted until they had concluded. Follow-up questions were asked when they brought up topics that the interviewer found interesting or considered meaningful for the adolescent to elaborate on. Furthermore, because the qualitative study was intended to complement the quantitative study, we perceive the length and depth of the interviews as sufficient.

One interview was conducted with 2 adolescents together, based on the participants’ own desires. Since they stated it would increase their sense of security during the interviews, the request was accepted. However, we cannot exclude the possibility that the presence of another adolescent influenced the answers the adolescents provided.

Conclusions

Mental health apps can be a useful resource for adolescents, and several apps geared toward adolescents have been developed. Our findings show that there is a lack of knowledge about the existence of mental health apps among adolescents. A large proportion of adolescents expressed that they would use these apps if they knew they were available. This study shows that apps directed toward adolescents should be easily accessible, free of charge, and provide easily understandable information. Adolescents also emphasized the significance that apps should be developed by credible sources or institutions, offer a choice between human and web-based support, and enable users to remain anonymous while seeking help. Future development of mental health apps should take these considerations into account.
Acknowledgments

The authors would like to thank the adolescents who participated in this study, both those who answered the web-based questionnaire and those who were interviewed. Further, the authors would like to thank Eirik Fjukstad and Annelene Moberg for helping with the recruitment and promotion of the study to adolescents in youth clubs and adolescent health centers in Tromsø. The publication charges for this paper have been funded by a grant from the publication fund of UiT The Arctic University of Norway.

Data Availability

The data sets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors provided substantial contributions to this study. HH and MR analyzed the qualitative data and HK and SK analyzed the quantitative data. HH wrote the first draft of the manuscript. HK, MR, SK read, edited, and approved the final manuscript.

Conflicts of Interest

None declared.

References


©Helene Høgsdal, Henriette Kyrrestad, Marte Rye, Sabine Kaiser. Originally published in JMIR Formative Research (https://formative.jmir.org), 15.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Original Paper

The Impact of Social Media Use on Mental Health and Family Functioning Within Web-Based Communities in Saudi Arabia: Ethnographic Correlational Study

Bdour Alwuqaysi1, MA; Alfie Abdul-Rahman1, PhD; Rita Borgo1, PhD

King's College London, London, United Kingdom

Corresponding Author:
Bdour Alwuqaysi, MA
King's College London
155 Wandsworth Road
Apt 3004, Sky Gardens
London, SW8 2FZ
United Kingdom
Phone: 44 07470334344
Email: bdour.alwuqaysi@hotmail.com

Abstract

Background: In recent years, increasing numbers of parents, activists, and decision-makers have raised concerns about the potential adverse effects of social media use on both mental health and family functioning. Although some studies have indicated associations between social media use and negative mental health outcomes, others have found no evidence of mental health harm.

Objective: This correlation study investigated the interplay between social media use, mental health, and family functioning. Analyzing data from 314 users, this study explores diverse mental health outcomes. The study places particular emphasis on the Saudi Arabian sample, providing valuable insights into the cultural context and shedding light on the specific dynamics of social media's impact on mental well-being and family dynamics in this demographic context.

Methods: We collected data through a subsection of an anonymous web-based survey titled “The Effect of COVID-19 on Social Media Usage, Mental Health, and Family Functioning.” The survey was distributed through diverse web-based platforms in Saudi Arabia, emphasizing the Saudi sample. The participants indicated their social media accounts and estimated their daily use. Mental health was assessed using the General Health Questionnaire and family functioning was evaluated using the Family Assessment Device Questionnaire. In addition, 6 mental health conditions (anxiety, self-esteem, depression, body dysmorphia, social media addiction, and eating disorders) were self-reported by participants.

Results: The study demonstrates a pattern of frequent social media use, with a significant portion dedicating 3-5 hours daily for web-based activities, and most of the sample accessed platforms multiple times a day. Despite concerns about social media addiction and perceived unhealthiness, participants cited staying connected with friends and family as their primary motivation for social media use. WhatsApp was perceived as the most positively impactful, whereas TikTok was considered the most negative for our Saudi sample. YouTube, Instagram, and Snapchat users reported poorer mental health compared with nonusers of these platforms. Mental health effects encompassed anxiety and addiction, with age and gender emerging as significant factors. Associations between social media use and family functioning were evident, with higher social media quartiles correlating with a greater likelihood of mental health and unhealthy family functioning. Logistic regression identified age and gender as factors linked to affected mental health, particularly noting that female participants aged 25-34 years were found to be more susceptible to affected mental health. In addition, multivariable analysis identified age and social media use quartiles as factors associated with poor family functioning.

Conclusions: This study examined how social media affects mental health and family functioning in Saudi Arabia. These findings underscore the need for culturally tailored interventions to address these challenges, considering diverse demographic needs. Recognizing these nuances can guide the development of interventions to promote digital well-being, acknowledging the importance of familial connections in Saudi society.

(JMIR Form Res 2024;8:e44923) doi:10.2196/44923

https://formative.jmir.org/2024/1/e44923

JMIR Form Res 2024 | vol. 8 | e44923 | p.198
(page number not for citation purposes)
KEYWORDS

social media use; mental health; family functioning

Introduction

Individual’s lives worldwide are now mostly impacted by social media [1]. By enabling individuals to retain offline connections and provide a more welcoming setting for emotional self-disclosure and help-seeking, social media might have a positive influence on their lives [2]. However, there have been rising worries about the possible detrimental impact of social media on individual mental health among certain scholars, governmental organizations, and the public [3]. Prior research has related social media use to multiple mental health conditions including body dissatisfaction, eating disorders, depressive symptoms, and social anxiety [4]. Facebook users, for instance, report higher body dissatisfaction than nonusers, according to a study involving male and female adults [5]. Another example of female adults revealed that although there was no difference between Instagram users and nonusers in terms of body dissatisfaction, Instagram users reported greater body monitoring than nonusers [6,7]. However, each of these works studied these conditions separately, and the combination of these conditions and how they are impacted by social media use has been overlooked. Given this, it becomes evident that there is a pressing need to conduct more in-depth investigations into the specific factors related to social media use that contribute to the development or exacerbation of these mental health concerns.

Family functioning pertains to how well family members fulfill their responsibilities and navigate life challenges [8]. It involves their interactions, growth, and responses to external influences [9]. Research highlights the need to investigate the potential risks posed by mobile device use to family functioning [10,11]. This suggests that families with flexibility and limited mobile device use may experience better cohesion and functioning, emphasizing the necessity for a more in-depth examination of the relationship between family functioning and social media use. Excessive social media use may lead to decreased emotional well-being, which negatively affects relationships [12]. For example, recent Indonesian research connects social media addiction to mild depression in university students, emphasizing the importance of bracing mental health by promoting family relationships and religiosity while addressing social media overuse [13]. Another study of mental health outpatients suggested a potential association between family functioning and psychological distress [14]. Although research has explored the link between social media and factors such as social connectedness, friendship quality, emotional well-being, and interpersonal relationships [15,16], there is a research gap regarding the impact of social media use on family functioning. Thus, it is imperative to conduct comprehensive research on family functioning and mental health, while analyzing their correlation with social media use.

In this paper, we looked at the relationships between social media use, mental health, and family functioning. Specifically, we examined various mental health conditions such as anxiety, self-esteem, depression, body dysmorphia, addiction to social media, and eating disorders. We selected Saudi Arabia as our sample for this study because of the increasing rates of internet penetration and social media use (90%) [17]. This high prevalence provides a valuable opportunity to examine the potential impact of social media on mental health and family functioning in this specific cultural context. Family is a fundamental unit in Saudi Arabian culture, and understanding the dynamics altered by social media is essential to comprehending broader societal changes. Investigating this aspect will not only enrich the existing literature, but also provide valuable insights into the evolving role of technology in shaping familial relationships and dynamics within the Saudi Arabian context. Thus, Saudi Arabia offers a unique setting characterized by a blend of traditional values, rapid modernization, and an evolving digital landscape. These factors create an intriguing environment in which to investigate the effects of social media on mental health and family dynamics.

Two studies on social media use in Saudi Arabia’s web-based community have been conducted, one highlighting the mental health problems associated with social media use [3] and the influence of social media on food consumption among individuals [18]. To our knowledge, there is no research on how social media use may affect family functioning in Saudi Arabia. Thus, our study aimed to provide insights tailored to this cultural context. We aimed to investigate the relationship between social media use and mental health outcomes, with a particular focus on age and gender differences. Gender plays a significant role in this context, and research indicates that women are more likely than men to experience these mental health issues and they also tend to use social media platforms more frequently than their male counterparts [19,20].

Methods

Procedures

The data collected in this study were gathered through a subsection of an anonymous web-based survey titled “The Effect of COVID-19 on Social Media Usage, Mental Health, and Family Functioning.” The survey was designed based on a review of previous studies and surveys on pandemic effects [21-24]. Recruitment for this survey used diverse web-based social media platforms such as Facebook and Twitter. The survey was distributed to participants in Saudi Arabia with a clear emphasis on targeting a Saudi sample. We obtained informed consent from all participants before they completed the survey. The survey was designed to be anonymous and voluntary. Participants were given the option to answer the survey in English or Arabic. The survey targeted social media users who were at least 18 years old, and included sections on demographics, COVID-19, social media use, and well-being. Refer to Multimedia Appendix 1 for more information on the survey.

Participants were asked to choose which of the following prominent social networking sites they had an account on: Instagram, Facebook, Twitter, Snapchat, Google+, Vine, Tumblr, Pinterest, YouTube, and others. Participants were also
provided with the option to specify “Other” platforms or “I do not have any social media accounts.” Participants were asked to estimate how much time they would spend on each platform during a typical day of browsing. No time (0), <5 minutes (1), 5-15 minutes (2), 15-30 minutes (3), 30 minutes -1 hour (4), 1-2 hours (5), 4-6 hours (7), 6-8 hours (8), 8-10 hours (9), or >10 hours (10). Participants were asked to respond to a variety of questions regarding the activities they engaged in on social media in general (not on specific platforms). Only participants who indicated that they had at least one social media account were presented with questions on social media activities. The web-based survey used in this study, designed based on a review of previous studies and surveys on the pandemic, can serve as a model for future research on the effects of social media use on mental health and family functioning.

Measures

Mental Health Status

We used the General Health Questionnaire-12 (GHQ-12), a well-established 12-item self-report assessment tool, to assess the mental health status of the survey participants. The Likert Scale was used to score all 12 questions in the GHQ-12, generating 3 distinct statistical indicators: typical, suggestive evidence of distress, and severe problems with psychological distress. This tool is widely used to assess psychological distress and mental well-being. The validity and reliability of this scale have been demonstrated [25-27].

Family Functioning Status

We used the Family Assessment Device Questionnaire (FAD) to estimate participants’ family functioning status. FAD is a self-reported scale specifically designed to provide insights into the overall dynamics and functionality within a family. It assesses family relationships and identifies areas of potential dysfunction by adhering to the McMaster Model of Family Functioning. Within the scope of this study, participants were presented with the general functioning scale of the FAD, comprising 12 questions and yielding 4 distinct statistical outputs: healthy, almost healthy, almost unhealthy, and unhealthy. FAD was chosen for this study because it best suited the study objectives and demonstrated its validity and reliability [28-30].

Mental Health Conditions

We present 6 mental health conditions—anxiety, self-esteem, depression, body dysmorphia, addiction to social media, and eating disorders—collected through direct survey questions where participants self-reported whether social media affected those conditions. The inclusion of these conditions in our study is justified based on their prevalence, established links to social media use, public health importance, diverse impacts on mental health, and practical implications for interventions and policies [31-34]. Notably, these issues are dominant, with a high incidence of social media addiction among individuals experiencing these mental health challenges [35].

Pilot Testing

To assess the initial survey, 4 participants participated in a pilot test. User feedback was collected to identify potential problems. The study had improved readability and validity because of this iterative process.

Statistical Analysis

Descriptive statistics were reported as numbers and percentages for categorical variables. The mean and SD are reported for the numerical values. A score was calculated based on the frequency of social media access and the average time spent on social media. The participants were classified based on their scores into 4 quartiles (Q1-Q4). Participants in the first quartile had the lowest social media use, whereas those in the fourth quartile had the highest social media use. To calculate the mental health score for the GHQ-12, we summed the assigned values (0-3) for each response, with higher scores indicating a greater likelihood of mental health issues. For family functioning using the FAD, we assigned values (1-4) to each response, with higher scores indicating a greater likelihood of unhealthy family functioning. To assess self-reported mental health conditions, including anxiety, self-esteem, and depression, we analyzed respondent’s answers to identify correlations with their social media scores. A chi-square test was performed to determine the association between social media use, mental health, and family functioning scores. Logistic regression was performed to identify the factors associated with mental health and family functioning. SPSS 28 (IBM Corp) was used for the analysis, and statistical significance set at $P<.05$ is considered statistically significant.

Recruitment

A total of 314 social media users who participated in this study were surveyed between the periods of 2021 to 2023 across 2 rounds to validate the results. Furthermore, 74.5% (n=234) were female, and 24.2% (n=76) were male. Most participants (n=293, 93.3%) were from Saudi Arabia, whereas the rest (n=21, 6.7%) represented other nationalities because of the nature of web-based sampling. The Saudi sample serves as an interesting case study for investigating the impact of social media on mental health and family functioning. First, Saudi Arabia is a highly conservative society that is undergoing rapid modernization, with social media playing a significant role in this transformation. According to recent statistics, 29.10 million social media users in Saudi Arabia access it through their mobile devices [17]. Second, there is a lack of research on the effects of social media on mental health and family functioning in Saudi Arabia. Finally, given that Saudi Arabia is a highly collectivist society, family dynamics play a significant role in shaping individual behaviors and attitudes, making it an ideal context to explore the interplay between social media use and family functioning. The largest group of participants was aged 35-44 years old (n=80, 25.5%), followed by 55-64 years (n=75, 23.9%). Regarding psychological and medical conditions, most respondents did not report having any psychological (n=271, 86.3%) or medical condition (n=216, 68.8%). Regarding the respondents’ educational background, the highest reported level of education was a bachelor’s degree (n=138, 43.9%), followed closely by a doctorate (n=81, 25.8%). A small proportion of respondents reported having a master’s degree (n=60, 19.1%), whereas a minority reported having a high school degree (n=26, 8.3%). Only a small percentage of the respondents reported having no formal education (n=9, 2.9%). Regarding participants’
race, most respondents identified as Arab (n=284, 90.4%), whereas a minority identified as non-Arab (n=30, 9.6%). Table S1 in Multimedia Appendix 2 provide more details on the sample demographics.

**Ethical Considerations**

The Effect of COVID-19 on Social Media Usage, Mental Health, and Family Functioning survey was ethically approved by King’s College London ethics committee LRS-19/20-19717.

**Results**

The purpose of this report was to summarize the results of a survey on the perceived impact of social media platforms on mental health.

**Social Media Use**

The data show that most respondents accessed social media frequently, with (n=115, 36.6%) reporting that they go on social media every couple of hours. When asked about the amount of time they spent on social media per day, the most common response was 3-5 hours (n=86, 27.4%), followed by 1-2 hours (n=58, 18.5%). When asked about the duration of their social media sessions, most respondents reported spending approximately 15 minutes or less logged in (n=129, 41.1%). A significant majority of respondents (n=238, 75.5%) reported that they did not feel it was healthy to spend much time on the internet. Most respondents (n=267, 85.03%) reported accessing social media in the evening, whereas 25.48% (n=80) reported accessing it at midnight. When asked about their addiction to social media, 41.1% (n=129) of the respondents reported feeling addicted. Table 1 provides more details on the social media use of the participants. The most common reason for using social media was to keep in touch with friends and family (n=243, 77.39%). Other reasons included inspiration (n=160, 50.96%), browsing or wasting time (n=135, 42.99%), and entertainment (n=130, 41.4%). Around 29.94% (n=94) of the participants reported using social media for work or business purposes, and only 5.1% (n=16) reported using social media for dating or romantic purposes. Table 2 elaborates the reasons for using social media.
<table>
<thead>
<tr>
<th>Table 1. Social media use among participants.</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>How often currently do you go on social media?</strong></td>
<td></td>
</tr>
<tr>
<td>Almost never or rarely</td>
<td>1 (0.3)</td>
</tr>
<tr>
<td>Just about every month</td>
<td>3 (1.0)</td>
</tr>
<tr>
<td>Every couple of days</td>
<td>10 (3.2)</td>
</tr>
<tr>
<td>Just about every day</td>
<td>83 (26.4)</td>
</tr>
<tr>
<td>Every couple of hours</td>
<td>115 (36.6)</td>
</tr>
<tr>
<td>Just about every hour</td>
<td>55 (17.5)</td>
</tr>
<tr>
<td>Every couple of minutes</td>
<td>47 (15.0)</td>
</tr>
<tr>
<td><strong>On average how much time do you think you spend on social media per day? (h)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;1</td>
<td>25 (8.0)</td>
</tr>
<tr>
<td>1-2</td>
<td>58 (18.5)</td>
</tr>
<tr>
<td>2-3</td>
<td>63 (20.1)</td>
</tr>
<tr>
<td>3-5</td>
<td>86 (27.4)</td>
</tr>
<tr>
<td>5-7</td>
<td>47 (15.0)</td>
</tr>
<tr>
<td>&gt;7</td>
<td>35 (11.1)</td>
</tr>
<tr>
<td><strong>Every time you log in to social media, on average how long do you spend logged in?</strong></td>
<td></td>
</tr>
<tr>
<td>About 15 min or less</td>
<td>129 (41.1)</td>
</tr>
<tr>
<td>About 30 min</td>
<td>83 (26.4)</td>
</tr>
<tr>
<td>About an hour</td>
<td>65 (20.7)</td>
</tr>
<tr>
<td>More than an hour</td>
<td>37 (11.8)</td>
</tr>
<tr>
<td><strong>Do you feel it is healthy to spend that much time online?</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>237 (75.5)</td>
</tr>
<tr>
<td>Yes</td>
<td>77 (24.5)</td>
</tr>
<tr>
<td><strong>When do you currently access social media? (multiple answers allowed)</strong></td>
<td></td>
</tr>
<tr>
<td>Morning (from 5 AM to 11:59 AM)</td>
<td>149 (47.45)</td>
</tr>
<tr>
<td>Afternoon (from 12 PM to 6 PM)</td>
<td>125 (39.81)</td>
</tr>
<tr>
<td>Evening (from 6 PM to 11:59 PM)</td>
<td>267 (85.03)</td>
</tr>
<tr>
<td>Midnight (exactly 12 AM to 4:59 AM)</td>
<td>80 (25.48)</td>
</tr>
<tr>
<td><strong>Do you consider yourself addicted to social media?</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>185 (58.9)</td>
</tr>
<tr>
<td>Yes</td>
<td>129 (41.1)</td>
</tr>
</tbody>
</table>
Table 2. Reasons for using social media.

<table>
<thead>
<tr>
<th>What do you use social media for? (multiple answers allowed)</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keeping in touch with friends and family</td>
<td>243 (77.39)</td>
</tr>
<tr>
<td>Event planning</td>
<td>49 (15.61)</td>
</tr>
<tr>
<td>Buying and selling</td>
<td>50 (15.92)</td>
</tr>
<tr>
<td>Inspiration</td>
<td>160 (50.96)</td>
</tr>
<tr>
<td>News about COVID-19 (coronavirus) pandemic</td>
<td>125 (39.81)</td>
</tr>
<tr>
<td>To make new friends</td>
<td>8 (2.55)</td>
</tr>
<tr>
<td>To find employment</td>
<td>13 (4.14)</td>
</tr>
<tr>
<td>To browse or time waste</td>
<td>135 (42.99)</td>
</tr>
<tr>
<td>To raise awareness</td>
<td>62 (19.75)</td>
</tr>
<tr>
<td>To provide support to others</td>
<td>58 (18.47)</td>
</tr>
<tr>
<td>To share your posts</td>
<td>66 (21.02)</td>
</tr>
<tr>
<td>To work</td>
<td>76 (24.20)</td>
</tr>
<tr>
<td>None of the above</td>
<td>3 (0.96)</td>
</tr>
</tbody>
</table>

Participants Attitudes

When asked if social media distracted them when they needed to be productive, 33.8% (n=106) of the respondents reported that it did. In contrast, 66.2% (n=208) of the respondents reported that social media does not distract them when they need to be productive. The data also revealed that a significant majority of respondents (n=238, 75.5%) did not care about how many people like or view their posts or pictures, whereas 24.5% (n=77) of respondents reported that they do care. When asked about cyberbullying on social media, 12.1% (n=38) of the respondents reported that they had been cyberbullied in some way, whereas 87.9% (n=276) reported that they had not. In terms of how social media affects self-esteem, only 30.9% (n=97) of the respondents reported feeling depressed when seeing posts about intriguing events in other people’s lives. When asked if they accept friend requests or followers from people they do not know to appear more popular, 84.7% (n=266) of the respondents reported that they did not, whereas 15.3% (n=48) reported that they did.

Social Media Impact via Self-Reporting

Participants were asked to rate each platform on a scale of 0 to 5, with 0 indicating that they did not use the platform and 5 indicating that it was the most positive social media platform in their opinion. After those who reported not using the platform, an average score was calculated based on each participant’s opinion, with a lower score indicating a negative impact and a higher score indicating a positive impact.

The platform perceived as having the most positive effect was WhatsApp (4.08), followed by Telegram (3.86), and Pinterest (3.85). The lowest score, indicating the most negative outcome, was observed for TikTok (1.98), followed by Snapchat (3.02). Table S2 in Multimedia Appendix 2 presents a comprehensive analysis of the perceived impact of each social media platform, as reported by the participants. The effects of individual platforms are shown in Figure 1. In addition, Figure 2 provides an overview of the mean scores indicating the perceived impact of social media platforms. Regarding self-reported mental health conditions, the most common effects were anxiety (n=217, 69.09%) and social media addiction (n=206, 65.45%). Other reported effects included depression (n=108, 34.55%), self-esteem (n=97, 30.91%), body dysmorphia (n=69, 21.82%), and eating disorders (n=46, 14.55%) as illustrated in Table 3. However, 25.45% (n=80) of the respondents who reported an impact stated that social media did not affect them. Finally, when asked about the emotions they experienced when using social networking sites, the most common responses were inspiration (n=238, 50.64%), motivation (n=208, 66.2%), and happiness (n=241, 54.8%). Other reported emotions included the fear of missing out (n=241, 29.3%), boosted self-esteem (n=208, 41.7%), jealousy (n=208, 34.5%), and rejection (n=208, 28.7%). Only 2.55% (n=8) of the respondents reported experiencing lower self-esteem when using networking sites. Table S3 in Multimedia Appendix 2 elaborates on social media use habits and their impact on participants.

Regarding the impact of social media on relationships with family members, 76.8% (n=241) of respondents reported that social media did not affect their relationships, whereas 23.2% (n=73) of respondents reported that it did. Among those who reported an impact, 15.3% (n=48) reported a positive effect, whereas 15.9% (n=50) reported a negative effect. When asked if they had a web-based persona, 91.1% (n=286) of the respondents reported that they did not, whereas only 8.9% (n=28) reported that they did.
Figure 1. The perceived effect of each social media platform as indicated by the participants.

Figure 2. Mean scores for the perceived effect of social media platforms.
Table 3. Social media impact.

<table>
<thead>
<tr>
<th>What impact has social media had on your mental health?</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety</td>
<td>38 (69.09)</td>
</tr>
<tr>
<td>Self-Esteem</td>
<td>17 (30.91)</td>
</tr>
<tr>
<td>Depression</td>
<td>19 (34.55)</td>
</tr>
<tr>
<td>Body dysmorphia</td>
<td>12 (21.82)</td>
</tr>
<tr>
<td>Addiction to social media</td>
<td>36 (65.45)</td>
</tr>
<tr>
<td>Eating disorder</td>
<td>8 (14.55)</td>
</tr>
<tr>
<td>It has not affected me</td>
<td>14 (25.45)</td>
</tr>
<tr>
<td>None of the above</td>
<td>22 (40.00)</td>
</tr>
</tbody>
</table>

Social Media Impact via Statistical Tools

Regarding mental health, most respondents (n=274, 87.3%) had a score that indicated mental health within the normal range, whereas (n=40, 12.7%) had a score that indicated affected mental health. For family functioning, most respondents (n=269, 85.7%) had a score that indicated healthy or almost healthy family functioning, whereas (n=45, 14.3%) had a score that indicated unhealthy or almost unhealthy family functioning. Table 4 shows the mental health and family functioning of the participants.

Statistical analysis using the chi-square test showed a statistically significant association between social media use and mental health ($P<.001$). Participants in the higher social media quartiles had a higher percentage of affected mental health (26.7% for the fourth quartile and 14.6% for the third quartile) as compared with participants in lower quartiles of social media use (9.1% in the first quartile and 4.8% in the second quartile). Table 5 shows the association between social media use and mental health. Statistical analysis using the chi-square test showed a statistically significant association between social media use and family functioning ($P<.001$). Participants in the higher social media quartiles had a higher percentage of unhealthy or almost unhealthy family functioning (30% for the fourth quartile and 14.6% for the third quartile) as compared with participants in the lower quartiles of social media use (9.1% for the first quartile and 8.3% for the second quartile). Table 6 illustrates the association between social media use and family functioning. Logistic regression was performed to identify the factors associated with mental health. Age and sex showed statistically significant results in the multivariate analysis. Female participants were more likely to have affected mental health as compared with male participants (odds ratio [OR] 4.69, 95% CI 1.42-15.49; $P=.01$). For age participants who were between 25 and 34 years were more likely to have affected mental health as compared with participants who were 18 to 24 years (OR 6.10, 95% CI 1.42-26.15; $P=.02$). Table 7 illustrates the factors associated with affected mental health. Logistic regression was applied to identify the factors associated with unhealthy or unhealthy family functioning. The age and social media use quartiles showed statistically significant differences. For gender, female participants were more likely to have unhealthy or almost unhealthy family functioning as compared with male participants (OR 3.32, 95% CI 1.17-9.46; $P=.02$). Regarding social media use quartiles, participants in the fourth quartile were more likely to have unhealthy or almost unhealthy family functioning as compared with participants in the first quartile (OR 4.22, 95% CI 1.45-12.31; $P=.008$). Table 8 illustrates the factors associated with unhealthy or almost unhealthy family functioning.

Table 4. Mental health and family function of participants.

<table>
<thead>
<tr>
<th>Mental health</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affected</td>
<td>40 (12.7)</td>
</tr>
<tr>
<td>Within normal range</td>
<td>274 (87.3)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Family functioning</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy or almost healthy</td>
<td>269 (85.7)</td>
</tr>
<tr>
<td>Unhealthy or almost unhealthy</td>
<td>45 (14.3)</td>
</tr>
</tbody>
</table>
**Table 5.** Association between social media use and mental health.

<table>
<thead>
<tr>
<th>Social media use quartiles</th>
<th>Mental health</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Within normal range, n (%)</td>
<td>Affected, n (%)</td>
<td>P value</td>
</tr>
<tr>
<td>Q1</td>
<td>80 (90.9)</td>
<td>8 (9.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Q2</td>
<td>80 (95.2)</td>
<td>4 (4.8)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Q3</td>
<td>70 (85.4)</td>
<td>12 (14.6)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Q4</td>
<td>44 (73.3)</td>
<td>16 (26.7)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

**Table 6.** Association between social media use and family functioning.

<table>
<thead>
<tr>
<th>Social media use quartiles</th>
<th>Family functioning</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Healthy or almost healthy, n (%)</td>
<td>Unhealthy or almost unhealthy, n (%)</td>
<td>P value</td>
</tr>
<tr>
<td>Q1</td>
<td>80 (90.9)</td>
<td>8 (9.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Q2</td>
<td>77 (91.7)</td>
<td>7 (8.3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Q3</td>
<td>70 (85.4)</td>
<td>12 (14.6)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Q4</td>
<td>42 (70)</td>
<td>18 (30.0)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
Table 7. Factors associated with affected mental health.

<table>
<thead>
<tr>
<th></th>
<th>Univariate</th>
<th></th>
<th></th>
<th>Multivariable</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds ratio (95% CI)</td>
<td>P value</td>
<td>Odds ratio (95% CI)</td>
<td>P value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (y)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>1 (N/A)</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25-34</td>
<td>4.77 (1.29-17.68)</td>
<td>.02</td>
<td>6.10 (1.42-26.15)</td>
<td>.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-44</td>
<td>1.65 (0.43-6.32)</td>
<td>.47</td>
<td>1.74 (0.42-7.27)</td>
<td>.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-54</td>
<td>0.97 (0.22-4.36)</td>
<td>.97</td>
<td>1.31 (0.26-6.66)</td>
<td>.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55-64</td>
<td>0.43 (0.08-2.25)</td>
<td>.32</td>
<td>0.39 (0.06-2.44)</td>
<td>.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65-74</td>
<td>0.00 (0.00)</td>
<td>&lt;.99</td>
<td>0.00 (0.00)</td>
<td>&lt;.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>3.27 (1.13-9.52)</td>
<td>.03</td>
<td>4.69 (1.42-15.49)</td>
<td>.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>0.00 (0.00)</td>
<td>&lt;.99</td>
<td>0.00 (0.00)</td>
<td>&lt;.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In Saudi Arabia</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outside Saudi Arabia</td>
<td>0.71 (0.16-3.15)</td>
<td>.65</td>
<td>0.96 (0.16-5.68)</td>
<td>.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level of education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High School</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>0.67 (0.23-1.99)</td>
<td>.47</td>
<td>0.34 (0.10-1.23)</td>
<td>.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master’s degree</td>
<td>0.55 (0.16-1.94)</td>
<td>.36</td>
<td>0.21 (0.04-1.00)</td>
<td>.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doctorate degree</td>
<td>0.34 (0.09-1.21)</td>
<td>.09</td>
<td>0.25 (0.06-1.15)</td>
<td>.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None of the above</td>
<td>2.10 (0.39-11.43)</td>
<td>.39</td>
<td>1.86 (0.21-16.73)</td>
<td>.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social media use quartiles</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td>0.28 (0.11-0.69)</td>
<td>.006</td>
<td>0.41 (0.11-1.60)</td>
<td>.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td>0.14 (0.04-0.44)</td>
<td>.001</td>
<td>1.47 (0.49-4.47)</td>
<td>.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td>0.47 (0.20-1.09)</td>
<td>.08</td>
<td>2.36 (0.78-7.16)</td>
<td>.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do you consider yourself addicted to social media?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.59 (0.30-1.15)</td>
<td>.12</td>
<td>0.83 (0.36-1.94)</td>
<td>.67</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*aN/A: not applicable.*
Table 8. Factors associated with unhealthy or almost unhealthy family functioning.

<table>
<thead>
<tr>
<th></th>
<th>Univariate</th>
<th></th>
<th></th>
<th>Multivariable</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds ratio (95% CI)</td>
<td>P value</td>
<td>Odds ratio (95% CI)</td>
<td>P value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (y)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>1 (N/A&lt;sup&gt;a&lt;/sup&gt;)</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25-34</td>
<td>1.24 (0.42-3.69)</td>
<td>.69</td>
<td>1.43 (0.40-5.09)</td>
<td>.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-44</td>
<td>0.82 (0.28-2.41)</td>
<td>.72</td>
<td>0.83 (0.25-2.77)</td>
<td>.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-54</td>
<td>0.86 (0.28-2.66)</td>
<td>.79</td>
<td>1.08 (0.29-4.07)</td>
<td>.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55-64</td>
<td>0.41 (0.12-1.37)</td>
<td>.15</td>
<td>0.39 (0.09-1.62)</td>
<td>.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65-74</td>
<td>0.00 (0.00)</td>
<td>&lt;.99</td>
<td>0.00 (0.00)</td>
<td>&lt;.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>2.84 (1.08-7.49)</td>
<td>.04</td>
<td>3.32 (1.17-9.46)</td>
<td>.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>4.73 (0.41-54.20)</td>
<td>.21</td>
<td>3.89 (0.29-52.68)</td>
<td>.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In Saudi Arabia</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outside Saudi Arabia</td>
<td>0.28 (0.04-2.16)</td>
<td>.22</td>
<td>0.46 (0.05-4.04)</td>
<td>.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level of education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High School</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>0.75 (0.26-2.22)</td>
<td>.61</td>
<td>0.51 (0.15-1.68)</td>
<td>.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master’s degree</td>
<td>0.30 (0.07-1.23)</td>
<td>.09</td>
<td>0.27 (0.06-1.26)</td>
<td>.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doctorate degree</td>
<td>0.73 (0.23-2.31)</td>
<td>.59</td>
<td>0.89 (0.24-3.27)</td>
<td>.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None of the above</td>
<td>2.10 (0.39-11.43)</td>
<td>.39</td>
<td>3.36 (0.45-24.99)</td>
<td>.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social media use quartiles</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td>0.91 (0.31-2.63)</td>
<td>.86</td>
<td>0.94 (0.30-2.93)</td>
<td>.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td>1.71 (0.66-4.43)</td>
<td>.27</td>
<td>1.89 (0.65-5.46)</td>
<td>.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td>4.29 (1.72-10.68)</td>
<td>.002</td>
<td>4.22 (1.45-12.31)</td>
<td>.008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do you consider yourself addicted to social media?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td>1.00 (N/A)</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.45 (0.77-2.73)</td>
<td>.25</td>
<td>0.72 (0.33-1.58)</td>
<td>.42</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>N/A: not applicable.

**Discussion**

**Principal Findings**

This study’s insights, grounded in a sample from Saudi Arabia, provide a culturally specific lens at the intersection of social media use and mental health. The prevalence of frequent social media access and reported durations align with the global trend [36], highlighting the pervasive nature of these platforms in the Saudi context. However, it is crucial to interpret these findings within the cultural framework of Saudi Arabia, where familial and social ties are of significant importance [37]. The primary motivation for social media use, namely staying connected with friends and family, resonates strongly with the cultural emphasis on community bonds in Saudi society. This underscores the integral role that social media plays in facilitating and maintaining relationships, which is a culturally significant function. The findings of our study suggest that social media is a popular and frequently used technology in Saudi Arabia, with a significant proportion of users expressing concern about their use habits. Unlike previous research, which analyzed social media platforms and their effect on mental health [38-41], our study applied an in-depth investigation across platforms to evaluate each platform’s impact on mental health. We first tested the effect of time spent on each platform by participants and found no particular association between time spent on various platforms and mental health. Despite small negative correlations between time spent on YouTube, Instagram, and Snapchat and body satisfaction and a small positive correlation between time spent on YouTube and depressive symptoms. Our findings...
imply that future research might benefit from changing attention from time spent generally perusing platforms to participant’s attitudes when engaging with social media platforms. The platform-specific analysis, with WhatsApp emerging as the most positively perceived and TikTok as the most negatively perceived, was influenced by cultural preferences and content norms in the Saudi context. Understanding these variations is essential for tailoring interventions and guidelines to align with the cultural values and sensitivities of the Saudi population.

Participants’ attitudes toward social media platforms may have affected their mental health. For instance, these platforms may be a hotspot for frequent and unjustified comparisons of appearances, which might be harmful to mental health. Performing more appearance comparisons with others and thinking that others are more attractive than you are on social media were both independent predictors of lower body satisfaction, more eating disorders, and higher levels of eating disorders. Our results support previous research on teenagers and adults, emphasizing the significance of appearance comparisons as a potential mechanism through which social media use may be detrimental to mental health [42,43]. Our results further imply that, despite their emphasis on physical appearance, these comparisons may have a detrimental effect on issues that are not just related to beauty, such as body satisfaction and eating disorders, but also on general mental health (such as depressive symptoms and anxiety). In our results, compared with men, women frequently paid greater attention to and regarded their beauty as a measure of their self-worth. Therefore, women may be less satisfied with their appearance and more depressed than men are. This is because women may engage in more frequent appearance comparisons on social media [44,45]. The identified effects on mental health, particularly anxiety and addiction, have cultural implications. Given the societal importance placed on mental well-being in Saudi Arabia, these findings underscore the need for targeted mental health awareness and support initiatives within a cultural context.

Logistic regression findings indicate that age and gender are factors associated with affected mental health and unhealthy family functioning. This is in line with a study that emphasized the significance of demographic factors when studying mental health in a Saudi sample [46]. According to our logistic regression findings, age and gender significantly influenced mental health and family functioning in the Saudi context. Female participants had a higher likelihood of experiencing mental health issues (OR 4.69, 95% CI 1.42-15.49; P=.01), emphasizing the need for gender-specific support. In addition, participants aged 25-34 years were more likely to face mental health challenges than those aged 18-24 years (OR 6.10, 95% CI 1.42-26.15; P=.02), recognizing the importance of age-targeted interventions. In terms of family functioning, female participants were more likely to report unhealthy dynamics (OR 3.32, 95% CI 1.17-9.46; P=.02), whereas older individuals in higher social media use quartiles were more likely to experience such challenges (OR 4.22, 95% CI 1.45-12.31; P=.008). Recognizing these age and gender dynamics is vital for tailoring mental health and family support strategies in Saudi Arabia. The statistical tools revealed associations between social media use and mental health, as well as family functioning, emphasizing the need for culturally informed strategies to address potential challenges.

The statistical analysis revealed significant associations between social media use, mental health, and family functioning within the Saudi Arabian context, underscoring the importance of culturally informed strategies. Higher social media quartiles exhibited a notable correlation with a greater likelihood of affecting mental health and unhealthy family functioning. These findings emphasize the nuanced interplay between web-based activities and individual well-being as well as the broader impact on familial relationships. Considering these associations, it is crucial to develop interventions and support mechanisms that are culturally sensitive and tailored to the unique sociocultural dynamics of Saudi Arabia. Recognizing the intricate relationship between social media use and mental health outcomes, along with its implications for family functioning, is the key to fostering digital well-being in this cultural context.

**Limitation and Implications**

It is important to consider the following limitations when interpreting the results of our study. More in-depth longitudinal studies are needed to explore the association between social media use and mental health over time. The sample’s specificity of the sample to Saudi Arabia’s demographic and cultural context may restrict the generalizability of the results to more diverse populations. To enhance the external validity, future research should aim for a broader and more representative sample that encompasses a range of cultural, socioeconomic, and demographic backgrounds. The study’s implications of this study are multifaceted and have significant relevance for the development of targeted interventions and public health initiatives in Saudi Arabia. First, the identified associations between social media use and mental health outcomes underscore the need for specific culturally sensitive interventions. Tailored mental health programs can address the distinct challenges faced by different demographic groups, such as female participants and individuals aged 25 to 34 years, who were found to be more susceptible to affected mental health. These interventions could include educational campaigns, support groups, and digital resources tailored to the cultural nuances of the Saudi context. Moreover, the observed link between social media use and family functioning emphasizes the interconnected nature of web-based behavior and familial relationships. Culturally informed strategies should not only address individual well-being, but also promote healthier family dynamics in the digital age. Public health campaigns can play a pivotal role in raising awareness of the potential impact of social media on family relationships and fostering open discussions within families and communities about responsible digital practices. This study not only contributes to the global discourse on social media, mental health, and family functioning but also offers nuanced insights specific to Saudi Arabia. Recognizing and understanding these cultural nuances are paramount for developing effective policies, educational programs, and support systems that promote positive mental health outcomes tailored to the sociocultural landscape of Saudi Arabia.
Conclusions
This study investigated the perceived impact of social media platforms on mental health and family functioning in a Saudi Arabian sample. The findings reveal important insights with implications for public health initiatives and targeted interventions. This study highlighted the observable association between social media use, mental health, and family functioning. Notably, age and gender have emerged as significant factors influencing mental health and unhealthy family functioning. This underscores the necessity for culturally sensitive strategies to address these identified challenges and tailor interventions to the specific needs of different demographic groups. Recognizing the nuanced associations observed in this study can inform the development of interventions that promote digital well-being, considering the crucial role of familial ties in the societal framework of Saudi Arabia.

Data Availability
The data sets generated during and analyzed during this study are available from the author upon reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Effect of COVID-19 on social media use, mental health, and family functioning survey.
[DOCX File, 3525 KB - formative_v8i1e44923_app1.docx ]

Multimedia Appendix 2
Statistical analysis tables.
[DOCX File, 33 KB - formative_v8i1e44923_app2.docx ]

References
12. Christensen SP. Social media use and its impact on relationships and emotions. Brigham Young University. 2018 Jun 1. URL: https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=7927&context=etd [accessed 2023-12-19]


23. Covid-19 social study - UK. University College London. URL: https://redcap.idhs.ucl.ac.uk/surveys/?s=TTXKND8JMK [accessed 2023-12-19]


Abbreviations

FAD: Family Assessment Device Questionnaire
GHQ-12: General Health Questionnaire-12
OR: odds ratio
Enabling Personalization for Digital Cognitive Stimulation to Support Communication With People With Dementia: Pilot Intervention Study as a Prelude to AI Development

Nick Hird¹, PhD; Tohmi Osaki², PhD; Samik Ghosh³, PhD; Sucheendra K Palaniappan³, PhD; Kiyoshi Maeda³, MD, PhD

¹Aikomi Ltd Co, Yokohama, Kanagawa, Japan
²Faculty of Rehabilitation, Kobe Gakuin University, Kobe, Japan
³Data Science and Engineering, SBX Corporation, Tokyo, Japan

Abstract

Background: Maintaining good communication and engagement between people with dementia and their caregivers is a major challenge in dementia care. Cognitive stimulation is a psychosocial intervention that supports communication and engagement, and several digital applications for cognitive stimulation have been developed. Personalization is an important factor for obtaining sustainable benefits, but the time and effort required to personalize and optimize applications often makes them difficult for routine use by nonspecialist caregivers and families. Although artificial intelligence (AI) has great potential to support automation of the personalization process, its use is largely unexplored because of the lack of suitable data from which to develop and train machine learning models.

Objective: This pilot study aims to evaluate a digital application called Aikomi in Japanese care homes for its potential to (1) create and deliver personalized cognitive stimulation programs to promote communication and engagement between people with dementia and usual care staff and (2) capture meaningful personalized data suitable for the development of AI systems.

Methods: A modular technology platform was developed and used to create personalized programs for 15 people with dementia living in 4 residential care facilities in Japan with the cooperation of a family member or care staff. A single intervention with the program was conducted with the person with dementia together with a care staff member, and for some participants, smell stimulation was provided using selected smell sticks in conjunction with the digital program. All sessions were recorded using a video camera, and the combined personalized data obtained by the platform were analyzed.

Results: Most people with dementia (10/15, 67%) showed high levels of engagement (>40 on Engagement of a Person with Dementia Scale), and there were no incidences of negative reactions toward the programs. Care staff reported that some participants showed extended concentration and spontaneous communication while using Aikomi, which was not their usual behavior. Smell stimulation promoted engagement for some participants even when they were unable to identify the smell. No changes in well-being were observed following the intervention according to the Mental Function Impairment Scale. The level of response to each type of content in the stimulation program varied greatly according to the person with dementia, and personalized data captured by the Aikomi platform enabled understanding of correlations between stimulation content and responses for each participant.

Conclusions: This study suggests that the Aikomi digital application is acceptable for use by persons with dementia and care staff and may have the potential to promote communication and engagement. The platform captures personalized data, which can provide suitable input for machine learning. Further investigation of Aikomi will be conducted to develop AI systems and create personalized digital cognitive stimulation applications that can be easily used by nonspecialist caregivers.

(JMIR Form Res 2024;8:e51732) doi:10.2196/51732
KEYWORDS
dementia; digital technology; communication; engagement; cognitive stimulation; artificial intelligence; AI

Introduction

Background
The lack of effective drugs for dementia [1] means that, for the foreseeable future, high-quality care remains the best option to maintain quality of life (QOL) for persons with dementia. However, cognitive decline and behavior changes associated with dementia increase the complexity and difficulty of caregiving, often making it challenging for families and care staff [2]. In particular, responsive behaviors, known medically as behavioral and psychological symptoms of dementia (BPSD), are a range of neuropsychiatric disturbances that affect most persons with dementia and can greatly disrupt caregiving, causing both poor QOL for the person with dementia and physical and mental stress for their caregivers [3]. Communication and engagement between people with dementia and their caregivers lie at the heart of good-quality caregiving [4,5], which is usually provided in dyadic or triadic structures [6] formed by the person with dementia and professional care staff and family caregivers. Communication plays a key role in the successful functioning and quality of care relationships in these care structures as well as greatly influencing the well-being and QOL of everyone involved [7]. Unfortunately, the progression of dementia can significantly impair the communication process for both persons with dementia and their caregivers and lead to inadequate or 1-sided interactions. Overcoming communication issues requires skill, patience, and sensitivity on the part of caregivers, which further adds to the difficulties and stress of caregiving [8]. In addition, the lack of time and resources available for caregivers can deprioritize communication as an activity in itself and result in it being conducted while performing other care activities, which may not be sufficiently personal or meaningful to maintain the psychological well-being of the person with dementia [9].

Communication and engagement are also integral to person-centered care [10], which has been shown to support QOL and help manage responsive behaviors and is widely accepted as best practice in dementia care. Communication difficulties between persons with dementia and their caregivers can prevent adequate expression and understanding of needs, desires, and intentions for both [11], presenting a major barrier to implementing person-centered care [12] including in Japan [13]. Communication and engagement are also important for implementing most psychosocial interventions [14], which are widely used to support caregiving activities to maintain well-being and QOL and manage BPSD. Many psychosocial interventions are based on different types of cognitive stimulation activities, such as cognitive training [15], reality orientation [16], reminiscence therapy [17], multisensory stimulation [18], and music therapy [19]. Several psychosocial interventions have shown promising clinical evidence, in some cases comparable with drug therapies, especially when implemented at the individual level. One of the most well-validated psychosocial interventions is cognitive stimulation therapy (CST) [20], which is based on person-centered care and consists of systematic protocols that combine reminiscence therapy and reality orientation designed to promote enjoyable and meaningful activities for persons with dementia. Clinical studies in groups have demonstrated improved cognition and QOL for persons with dementia [21] and improved caregiver relationships when conducted at the individual level (individual CST; iCST) [22]. Culturally adapted CST protocols have been developed for >20 countries, including CST-J for Japan [23], although its adoption in Japanese care settings remains limited, especially at the individual level.

The lack of trained care staff and practical difficulties associated with the regular and consistent implementation of psychosocial interventions have generated considerable interest in the use of digital technologies [24,25], a trend that was accelerated by the COVID-19 pandemic [26]. Recently, several applications have become available to promote personalized communication and engagement between people with dementia and their caregivers [27], and personalization has been recognized as an important factor in obtaining sustainable benefits [28]. Digital storytelling is a promising approach based on the well-established life story book concept in reminiscence therapy, which uses a person’s own and other relevant content to create fully individualized interventions [29]. In digital storytelling, the physical materials commonly used to facilitate life story book interventions, such as photographs, books, and memorable objects, are replaced with digital media, such as images, videos, and audio. In addition to engagement, digital storytelling aims to help the person tell their own story and has also been used outside dementia in other areas of mental health. Feasibility studies with digital storytelling applications have shown improvements in memory, QOL, and depression [30] as well as additional benefits from the use of digital media. Similar results have been observed in both Western and Asian contexts [31], indicating the wide potential of this approach. However, one of the difficulties of digital storytelling is that preparation requires digital skills, time, and effort [32], which presents a significant adoption barrier for many caregivers and families. The need to create personalized content is avoided through a digital iCST application [33] that uses a pool of precreated generic stimulation activities, including quizzes and games designed to promote engagement, that can be used according to the interests and preferences of the users. A feasibility study showed promising results [34]; however, the lack of personalized content and content diversity was identified as an issue for maintaining engagement and long-term use. Given the importance of personalization, reducing the time and effort required for tailoring applications is a key concern to facilitate broader adoption by nonexpert care staff and families. Data generated by digital technologies can be used to develop artificial intelligence (AI) systems to support caregiving for dementia [35], which can aid the personalization process. However, the development of personalization applications is still at an early stage, and the lack of high-quality personalized data sets related to communication and engagement for people with dementia hinders progress in this area. To address this, a prototype
application called Aikomi was developed to support communication with people with dementia while also capturing high-context personalized data, including behaviors, that would provide a suitable source for developing machine learning models to automate the personalization process. Such AI systems could reduce the personalization burden of digital cognitive stimulation applications and enable their use by non-specialist caregivers to support communication and engagement with people with dementia in real-world settings.

**Aims and Objectives**

The aim of this pilot study was to evaluate the Aikomi application in Japanese residential care homes for its potential to (1) create and deliver personalized cognitive stimulation programs and promote communication and engagement between people with dementia and their usual care staff and (2) capture meaningful personalized data suitable for the development of AI systems.

**Methods**

**Technology Development**

The Aikomi application was designed by a multidisciplinary team with expertise in occupational therapy (TO); clinical dementia care and psychiatry (KM); digital health (NH); engineering, data science, and machine learning (SG and SKP); the design process also received input from frontline care staff in Japan. The design goal was to create an application that could be routinely used by care staff and family members to promote communication with people with dementia living in both residential care and community settings and would not require expert knowledge, training, or long preparation times. The technology goals were to (1) construct the technology platform using a modular architecture with stand-alone applications for each key function to allow for “plug-and-play” integration of software and hardware applications and (2) incorporate an open connectivity platform to allow for convenient data flow between application modules and facilitate the development of machine learning applications. The connectivity platform used was Garuda (Garuda Alliance) [36], which is a community-driven open connectivity platform previously developed by one of the authors (SG). The modular design provides flexibility to develop fit-for-purpose applications as well as “future-proof” the platform to incorporate other technologies and data sources to enhance optimization and scaling of the personalization process.

A 6-step workflow for using the application was designed through informal discussions with professional care staff and families (Figure 1). The first step is to conduct an interview with families or care staff to obtain information about the person with dementia, including their life history, hobbies, interests, preferences, and abilities, which are used to create a standardized personal profile. On the basis of this personal profile, relevant digital media content is selected or created, and if available, the person’s private content obtained from the family is digitized (as necessary) and uploaded to the system. Next, relevant content is compiled into audiovisual stimulation programs, which are displayed as the intervention. Finally, the behavioral responses of the person with dementia during the intervention are recorded and analyzed to adapt and optimize the content and stimulation programs for the next intervention. To minimize the burden of the personalization process on families and care staff, their participation is limited to the initial interview and the intervention itself. The remaining personalization processes were conducted manually by the research team for this pilot study and will be automated in the future.

The software for all the components (modules) was developed as an independent web application whose integration was facilitated by the Garuda connectivity platform. The content management system (CMS) is a repository for generic digital media content, such as images, videos, and audio files. It allows users to upload, create, edit, and store content that can be used by all persons. A separate private CMS module performs the same function as the CMS for the private content of the person, but its use is restricted to this person only. The simulator module is a function to create stimulation elements (called STIMs) from the content in the CMS and private CMS. A STIM consists of a short audiovisual sequence created from combinations of image, audio, and video data and is the building block used to create stimulation programs. The user has complete flexibility to create and select STIMs and compile them in any sequence order to generate the stimulation program. The personal CMS and simulator modules operate at the level of each individual and are not accessible to other users and caregivers. The simulator module also includes a function to create and edit a standardized personal profile of the person. The home and control modules are both used to conduct the intervention. The home module displays the stimulation programs viewed by the person with dementia, and the control module is used by the caregiver to select and control the display of the stimulation program on the home module during the intervention. The control module has functions to pause the program to talk or go back to a previous STIM or forward to a new STIM depending on the response of the person with dementia. A web meeting function was integrated into the home and control modules to allow for the remote use of the Aikomi application, but it was not used in this pilot study, in which only in-person sessions were conducted. The design of the functions and user interfaces for the home and control modules was conducted in collaboration with care staff primarily for use with tablets, but they can also be used on a PC. An example of the home and control modules during use is shown in Figure 2.

The response dashboard module enables the storage, review, and data analysis of video recordings of the person with dementia during the intervention. It was intended for video recordings to be made using the camera on the home tablet; however, this function was not fully operational at the time of the pilot study. Hence, a remote camera was used to record the interventions, and the data were subsequently uploaded to the response dashboard. Each module was a separate web application integrated with the Garuda connectivity platform. The tablets used in the pilot study were installed with SIM cards to avoid the need to use the local network at the residential care home, which can sometimes be unreliable.
Figure 1. Aikomi workflow and system modules.

Aikomi workflow
- Interview with family or caregiver
- Create personal profile for user

Processes to be automated:
- Upload, select, create contents
- View and analyze user behavioral response to program

Manual processes:
- Compile contents into cognitive stimulation program
- Display program to user

Aikomi system modules
- Content management system: upload, create, edit and store generic digital media contents
- Home: display stimulation programs to person with dementia
- Private content management system: upload, create, edit, and store a user’s own digital media contents
- Control: caregiver operations to select and play stimulation programs
- Simulator: create, edit, store, and schedule stimulation elements and programs
- Response dashboard: record behavioral responses of person with dementia to program

Figure 2. Home and control module display.

Home module (viewed by person with dementia)

Control module (used by caregivers to conduct intervention)
Pilot Study

Study Design and Participants

The study was conducted using Aikomi as a single intervention lasting approximately 30 minutes with 15 people with dementia living in 4 urban residential care homes in Japan and their usual care staff. The participants were nominated by the care managers at each facility based on the following selection criteria: (1) a diagnosis of dementia, (2) displaying negative BPSD such as anxiety or apathy, (3) no hearing or visual difficulties that would prevent using a tablet, and (4) agreement from their family to participate in the study. Participants with a diagnosis of frontotemporal dementia or who had other mental conditions were excluded from this study. All the care staff members who participated in this study were qualified professionals at the residential facilities.

Procedure

The protocol and timeline for the pilot study are shown in Figure 3.

Participant Selection

Care managers at participating care homes explained the details of Aikomi and the pilot study to the families of candidate persons with dementia using materials provided by the research team. After the explanation, written consent was obtained from the families who agreed to participate in the study.

Program Preparation

The research team conducted in-person or telephone interviews with family members to obtain information about the person’s life story, interests, and preferences as well as other topics that the family thought could be meaningful. In addition, the families were asked to provide any suitable family photos if they had them. In one case, a family member was not available for interview, and it was instead conducted with the care manager at the facility. The obtained information was used to create a profile for each person, and relevant digital content was selected and used to create STIMs that corresponded to items in the profile, such as hometown, childhood, family, work, life and cultural events, hobbies, sports, travel, pets, and music. The duration of each STIM ranged from 30 seconds to 3 minutes, and each stimulation program was created by compiling 10 to 20 STIMs in a sequential order expected to be easy for the person with dementia to follow. For this pilot study, all non–family-derived content was obtained by the research team from publicly available sources, and the programs were prepared by the research team within approximately 2 weeks following the family interview.

Intervention

The intervention was conducted as a 1:1 session with the person with dementia and the care staff member seated next to each other at a table in a quiet area of the care home. The home tablet was placed at a comfortable viewing distance for the person with dementia and the care staff member. The control tablet was operated by a research team member seated on the opposite side of the table. A camera was placed at an appropriate position to record the behavioral responses of the person with dementia and care staff member during the intervention. As the application was to be operated by the research team, the care staff were only given a brief overview of the device, and a rehearsal was conducted before the session to familiarize them with the intervention conditions. In addition, to allow the person with dementia to lead responses to the stimulation program as much as possible, the care staff were requested to adopt a passive role during the intervention and respond appropriately according to the person's behavior and mood, although prompting according to their own judgment was permitted. The stimulation program was started as soon as the participant was seated and appeared comfortable, and apart from initial greetings, the research team member avoided direct interaction with the person with dementia during the session unless actively addressed by them. The intervention began with a few general reminiscence STIMs to act as a “warm up” and accustom the person with dementia to the device, after which STIMs related to personalized topics were played. The selection of STIMs was adjusted according to the mood and responses of the participant, and if a participant showed good engagement with a particular theme, the STIM was repeated or a related STIM was shown. In cases in which
the participant showed little engagement, the STIM was stopped and a STIM of a different topic was shown. The target time for each intervention was 30 minutes but was shortened if the participant appeared tired or uninterested, and the STIM was changed immediately if the participant appeared uncomfortable. The intervention was concluded by showing STIMs such as nature or landscapes as a “cool down” period, and if the participant was agreeable, smell stimulation was conducted using 5 synthetic smells that were embedded on paper strips (such as those used for sampling perfumes) supplied by a commercial smell product company [37]. The decision to use the smell sticks was made after asking each person with dementia and the care staff member after completion of the main stimulation program. The 5 smells used for the evaluation were chocolate, miso (fermented soy paste cooking ingredient), grass, earth, and soap. A smell stick was provided to both the participant and the caregiver without revealing the identity of the smell, and at the same time, a STIM related to the smell was displayed on the home module, for example, showing a bar of a famous brand of chocolate for the chocolate smell. Following the first smell, the remaining smell sticks were similarly used individually according to the participants’ interest in continuing.

Evaluation Scales

Baseline

BPSD assessment of the participants was conducted during the month before the intervention by the appropriate care staff using the Neuropsychiatric Inventory–Nursing Home version (NPI-NH) [38]. Cognitive ability was recorded using the most recent cognitive assessment conducted at the care home using one of the following cognitive scales: Mini-Mental State Examination (MMSE) [39], revised Hasegawa Dementia Scale (HDS-R) [40], or Nishimura Mental State Scale for the Elderly (NM) [41]. In total, 3 different cognitive scales were used because of the different types of cognitive tests routinely used at each participating care home.

Intervention

Engagement during the intervention was measured using a Japanese translation of the Engagement of a Person with Dementia Scale (EPWDS) [42] prepared by the research team. The EPWDS is a 10-item assessment measuring positive and negative engagement in 5 domains—affect, visual, verbal, behavioral, and social—and was administered soon after the intervention by the attending care staff member, who was also asked to provide their comments on the intervention. Psychological well-being was measured using the 6 motivational and emotional dysfunction items of the Mental Function Impairment Scale (MENFIS) [43]. The MENFIS assessment was conducted at the care home by a care staff member on 3 consecutive days spanning the intervention: the day before the intervention (day 1), on the day of the intervention a few hours after it was conducted (day 2), and the day after the intervention (day 3). The care staff member was also asked to provide their written comments for each assessment. All 3 MENFIS assessments were conducted by the same care staff member, who was not always the same care staff member who was present at the intervention. As the study was intended as a preliminary pilot study, the EPWDS and MENFIS scores were only used as a guide for the qualitative assessment of the intervention.

Ethical Considerations

Ethics approval was obtained from the Kobe Gakuin University Human Research Ethics Committee (sourin 18-14). Written approval to conduct the study was obtained from the directors of each of the participating care facilities as well as the families of all the persons with dementia participating in the study.

Results

Overview

The characteristics of the 15 participants enrolled in the study are shown in Table 1. A total of 5 (33%) people did not meet all the selection criteria: 2 (13%) were “suspected” to have Alzheimer disease but had not received a formal diagnosis of dementia, and 3 (20%) did not show BPSD as defined by the NPI-NH. However, it was decided to include all 5 in the pilot study as the primary purpose was to evaluate Aikomi for its acceptability and communication effects on persons with dementia, and there was agreement to proceed from the care staff and families. The people who did show BPSD were reported to have a wide range of symptoms.
Table 1. Baseline characteristics of persons with dementia in the pilot study.

<table>
<thead>
<tr>
<th>Person</th>
<th>Care home</th>
<th>Sex</th>
<th>Age (y)</th>
<th>Dementia type</th>
<th>Cognitive scores</th>
<th>BPSD</th>
<th>Care level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HDS-R</td>
<td>MMSE</td>
<td>NM</td>
</tr>
<tr>
<td>1</td>
<td>A</td>
<td>Male</td>
<td>93</td>
<td>AD</td>
<td>8</td>
<td>17</td>
<td>ND</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>Female</td>
<td>89</td>
<td>AD</td>
<td>7</td>
<td>11</td>
<td>ND</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>Female</td>
<td>86</td>
<td>AD</td>
<td>2</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>Female</td>
<td>86</td>
<td>AD</td>
<td>ND</td>
<td>21</td>
<td>ND</td>
</tr>
<tr>
<td>5</td>
<td>B</td>
<td>Male</td>
<td>79</td>
<td>AD</td>
<td>ND</td>
<td>24</td>
<td>ND</td>
</tr>
<tr>
<td>6</td>
<td>B</td>
<td>Male</td>
<td>76</td>
<td>AD</td>
<td>ND</td>
<td>8</td>
<td>ND</td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>Female</td>
<td>87</td>
<td>AD</td>
<td>ND</td>
<td>17</td>
<td>ND</td>
</tr>
<tr>
<td>8</td>
<td>B</td>
<td>Female</td>
<td>84</td>
<td>AD</td>
<td>ND</td>
<td>16</td>
<td>ND</td>
</tr>
<tr>
<td>9</td>
<td>B</td>
<td>Male</td>
<td>79</td>
<td>AD</td>
<td>ND</td>
<td>20</td>
<td>ND</td>
</tr>
<tr>
<td>10</td>
<td>B</td>
<td>Female</td>
<td>83</td>
<td>AD</td>
<td>ND</td>
<td>9</td>
<td>ND</td>
</tr>
<tr>
<td>11</td>
<td>B</td>
<td>Female</td>
<td>86</td>
<td>AD</td>
<td>ND</td>
<td>15</td>
<td>ND</td>
</tr>
<tr>
<td>12</td>
<td>C</td>
<td>Female</td>
<td>90</td>
<td>AD</td>
<td>6</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>13</td>
<td>C</td>
<td>Female</td>
<td>86</td>
<td>AD</td>
<td>0</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>14</td>
<td>D</td>
<td>Female</td>
<td>77</td>
<td>ND</td>
<td>ND</td>
<td>6</td>
<td>ND</td>
</tr>
<tr>
<td>15</td>
<td>D</td>
<td>Female</td>
<td>90</td>
<td>ND</td>
<td>ND</td>
<td>25</td>
<td>ND</td>
</tr>
</tbody>
</table>

a: group home; B: residential nursing home; C and D: older adult rehabilitation facility.

b: Reported dementia diagnosis.

c: BPSD: behavioral and psychological symptoms of dementia.
d: Japanese long-term care insurance rating scale [44].
e: HDS-R: revised Hasegawa Dementia Scale [40].
f: MMSE: Mini-Mental State Examination [39].
g: NM: Nishimura Mental State Scale for the Elderly [41].
h: NPI-NH: Neuropsychiatric Inventory–Nursing Home version [38].
i: AD: Alzheimer disease.
j: ND: not determined.

Engagement

The responses of the persons with dementia during the intervention are shown in Table 2 and Multimedia Appendix 1. The duration of the intervention ranged from 15 to 38 minutes, most persons (13/15, 87%) showed strong positive responses to at least 1 of the STIMs, and none showed discomfort toward Aikomi or requested the session to be stopped. In total, 67% (10/15) of the participants had an EPWDS score of >40 (out of a maximum of 50), indicating both a high incidence of positive engagement and a low incidence of negative engagement. The remaining 33% (5/15) of the participants showed an EPWDS score of >30, which was due to low positive engagement scores (lack of engagement with the stimulation program) rather than the high incidence of negative responses such as anger, anxiety, or discomfort, which were not observed for any participant. Interestingly, person 14, who showed few positive responses during the intervention itself, spoke to thank the research team after the session was completed, which the care staff member said was highly unusual behavior for them.

The types of STIM topics and the responses they generated were analyzed from the video recordings and are shown in Table 3. The most common STIM topic to generate good engagement was family photos, which prompted self-initiated talk and the identification of persons they recognized. However, some participants (6/15, 40%) struggled to recognize their family members and even themselves. One person (person 3) showed no response, and person 4 responded most strongly to photographs of herself in early adult life rather than of her family. Music was also a popular STIM topic, prompting several participants to initiate singing and clapping to both traditional Japanese children’s songs (doyo) and Japanese popular music. The STIM of popular singers known to be liked by the persons with dementia often generated verbal dialogue with the care staff (persons 4, 6, 8, and 13). Japanese traditional arts, such as the tea ceremony, dance, and calligraphy, generated good responses from most women who had performed them (persons 2, 4, 7, 8, 11, and 13), and sports themes (baseball, sumo, and boxing) were popular for all the men (persons 1, 5, 6, and 9). Work-related themes induced mixed responses; however, STIMs related to actions performed during their work (eg, using a
Japanese typewriter, counting money, or preparing fish) generated responses from 33% (5/15) of the participants (persons 2, 6, 7, 8, and 15). Wartime navy service STIMs generated strong positive engagement from person 1 and prompted a detailed recollection of his experiences; this is described in more detail in case study 1.

Table 2. Engagement during the session with the Aikomi device.

<table>
<thead>
<tr>
<th>Person</th>
<th>EPWDS score a</th>
<th>Caregiver written comment provided after the session (translation from Japanese) b</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>46</td>
<td>He showed very spontaneous reaction to family and the warship.</td>
</tr>
<tr>
<td>2</td>
<td>33</td>
<td>She concentrated for volleyball and knitting but could not remember family faces or names. She has severe memory loss.</td>
</tr>
<tr>
<td>3</td>
<td>37</td>
<td>She quickly responded to music, but the response is similar to what the care staff can obtain using tablets.</td>
</tr>
<tr>
<td>4</td>
<td>41</td>
<td>She was able to concentrate due to the music, and was very focused for one singer. Usually, her concentration doesn’t last for 5 minutes, it was very unusual for her to maintain concentration for 30 minutes.</td>
</tr>
<tr>
<td>5</td>
<td>48</td>
<td>He turned to look at the tablet as soon as the images appeared and focused on talking about them, including in great detail about the movies. This is behavior not usually observed by the care staff.</td>
</tr>
<tr>
<td>6</td>
<td>34</td>
<td>Despite being a shy person, he was able to sing in front of everyone. He usually can’t recognize things, but was able to for some pictures. He was anxious because he couldn’t understand many things.</td>
</tr>
<tr>
<td>7</td>
<td>42</td>
<td>She did hand gestures to songs and Japanese dance and hula dance. She did not say much because she is a naturally reserved person, but she showed concentration and seemed excited.</td>
</tr>
<tr>
<td>8</td>
<td>39</td>
<td>She looked at the tablet and talked continuously but not related to the themes shown. She showed good responses to music and smell.</td>
</tr>
<tr>
<td>9</td>
<td>46</td>
<td>He talked about his mother and explained to us in detail about his old hobbies. He was anxious because he was able to detect the smells.</td>
</tr>
<tr>
<td>10</td>
<td>45</td>
<td>She showed good expression and mood. Usually she doesn’t continue laughing, and I think it was due to the continuous stimulation.</td>
</tr>
<tr>
<td>11</td>
<td>45</td>
<td>She clapped her hands and sang to the music, which is the same response as she regularly shows with karaoke.</td>
</tr>
<tr>
<td>12</td>
<td>32</td>
<td>She had pneumonia before the test, which reduced her will, and she only responded weakly to themes.</td>
</tr>
<tr>
<td>13</td>
<td>40</td>
<td>She is usually not a person who can show good concentration but was calm and concentrated during the test and could recall her memories.</td>
</tr>
<tr>
<td>14</td>
<td>40</td>
<td>She showed most interest in the old photos. At the end of the test, she smiled and said “thank you”.</td>
</tr>
<tr>
<td>15</td>
<td>47</td>
<td>She sang along to the music and looked nostalgic when watching the old photos.</td>
</tr>
</tbody>
</table>

aEPWDS: Engagement of a Person with Dementia Scale. Range of scores is 10-50, higher scores indicate higher level of positive engagement, lower scores indicate higher levels of disengagement or negative engagement [42].

bTranslation by the research team.
### Table 3. Engagement by stimulation element (STIM) topic.

<table>
<thead>
<tr>
<th>Person</th>
<th>Self-initiated talk(^b)</th>
<th>Prompted talk(^b)</th>
<th>Notes(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Family, navy service</td>
<td>Hometown, baseball, photography</td>
<td>Viewed war service STIM several times, adding new anecdotes each time. Responded strongly to family photos.</td>
</tr>
<tr>
<td>2</td>
<td>Volleyball, Japanese typewriter, kimono</td>
<td>Family, music, cooking, childhood</td>
<td>Husband of participant 1 but did not respond to family photos. Was good at volleyball. Used typewriter at work.</td>
</tr>
<tr>
<td>3</td>
<td>None</td>
<td>Music</td>
<td>Only responded to music (sang). Did not react to personal topics.</td>
</tr>
<tr>
<td>5</td>
<td>Archery, 100-km walk, food, pigeon racing, gardening, childhood</td>
<td>Family, school, life, pet, hot spring bath, hometown</td>
<td>Talked in detail about participating in 100-km walk. Said that gardening STIM showed incorrect way to grow orchids. Talked about difficult times during childhood.</td>
</tr>
<tr>
<td>6</td>
<td>Boat race, Japanese singer, sumo, food</td>
<td>Baseball, cooking fish, family, music, cultural event</td>
<td>Nervous at first but calmed down when the program started. Talked about working at a fish restaurant. Did not recognize many family members. Sang to music.</td>
</tr>
<tr>
<td>7</td>
<td>Family, Hawaii, and Japanese arts</td>
<td>Bank, counting money</td>
<td>Talked about cousin who lived in Hawaii. Remembered how to count money. Looked closely at calligraphy and flower arranging.</td>
</tr>
<tr>
<td>8</td>
<td>Yakitori (grilled chicken), dog, Japanese poetry</td>
<td>Japanese singer, family, music, hometown</td>
<td>Talked continuously but not related to STIM. Stopped talking to look at yakitori and Japanese poetry. Sang to music.</td>
</tr>
<tr>
<td>9</td>
<td>Family, movies, boxing, baseball</td>
<td>Detective novel, dog</td>
<td>Talked about family life and children and in great detail about films, naming actors and directors. Identified boxers and baseball players and talked in detail.</td>
</tr>
<tr>
<td>10</td>
<td>Watercolor painting, hill walking, athletics</td>
<td>Family, knitting, television drama, food</td>
<td>Said that she wanted to try watercolor painting and knitting. Named some mountains she had climbed. Said she was fast at running at school.</td>
</tr>
<tr>
<td>11</td>
<td>Family, hometown, music, tea ceremony</td>
<td>Schools, school sports, music, cooking</td>
<td>Could identify more family members on repeated viewing. Repeated name of hometown several times. Watched tea ceremony closely for several minutes. Sang to music.</td>
</tr>
<tr>
<td>12</td>
<td>None</td>
<td>None</td>
<td>Did not look at the screen for most of the session. Showed no reaction to any STIM.</td>
</tr>
<tr>
<td>13</td>
<td>Hometown, Japanese singer, <em>koto</em> (Japanese musical instrument), childhood</td>
<td>Kobe, childhood songs</td>
<td>Talked about growing up in Tokyo and sad life of Japanese singer. Watched <em>koto</em> playing closely for several minutes.</td>
</tr>
<tr>
<td>14</td>
<td>None</td>
<td>None</td>
<td>Looked continuously at the tablet but did not show any reaction to any STIM.</td>
</tr>
<tr>
<td>15</td>
<td>Family, travel, childhood, hotel work</td>
<td>Paper making, hot spring bath, music, son’s work</td>
<td>Talked about family and climbing Mount Fuji with her son. Recalled working at a hotel and her son’s company.</td>
</tr>
</tbody>
</table>

\(^{a}\)Talk initiated by the person with dementia without prompting.  
\(^{b}\)Person responded to prompting by the care staff member.  
\(^{c}\)On the basis of a video review by the research team.

**Smell**

A total of 60% (9/15) of the participants tried the smell stick stimulation in conjunction with paired audiovisual stimulation, and 13% (2/15) of the participants tried all 5 smell sticks. The responses are shown in Table 4. In most cases, the smell sticks led to pleasurable responses from both the participant and the care staff member, with more laughter than was observed during the audiovisual programs. Almost all the people with dementia (8/9, 89%) had difficulty explicitly identifying the smell; however, majority (6/9, 67%) were able to notice that the smells were different, and recognition of the smell identity did not appear to affect the engagement with the caregivers. Although it was not the intention, some participants perceived the smell program as a test, which may have caused some confusion (person 15) and prompted one person to say that she had lost her sense of smell (person 5). This shows that care must be exercised when using smell stimulation to reduce the risk of causing anxiety to users who may no longer have a sense of smell or be stressed by the inability to correctly identify the smells.
Table 4. Response of participants to smell sticks combined with audiovisual stimulation.

<table>
<thead>
<tr>
<th>Person</th>
<th>Smell stick used</th>
<th>Response&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First</td>
<td>Second</td>
</tr>
<tr>
<td>1</td>
<td>Chocolate</td>
<td>Soap</td>
</tr>
<tr>
<td>3</td>
<td>Chocolate</td>
<td>Miso</td>
</tr>
<tr>
<td>5</td>
<td>Chocolate</td>
<td>Soap</td>
</tr>
<tr>
<td>8</td>
<td>Chocolate</td>
<td>Soap</td>
</tr>
<tr>
<td>9</td>
<td>Chocolate</td>
<td>Soap</td>
</tr>
<tr>
<td>10</td>
<td>Chocolate</td>
<td>Soap</td>
</tr>
<tr>
<td>11</td>
<td>Chocolate</td>
<td>Soap</td>
</tr>
<tr>
<td>13</td>
<td>Chocolate</td>
<td>—</td>
</tr>
<tr>
<td>15</td>
<td>Chocolate</td>
<td>Soap</td>
</tr>
</tbody>
</table>

<sup>a</sup>On the basis of a video review by the research team.
<sup>b</sup>Not used.

Well-Being and Behavior Changes

The results of the MENFIS well-being assessment of the persons with dementia conducted by the care staff over the consecutive 3-day intervention period are shown in Table 5. In total, 13% (2/15) of the participants showed improvement (reduction in MENFIS score) on the third day compared with the first, and 7% (1/15) of the participants (person 13) showed a reduction only on the second day (after the intervention). All other participants showed similar or identical MENFIS scores for all 3 days except for person 12, who showed worsening over the 3 days, which was attributed to her catching a cold during the period of the intervention. The data for persons 14 and 15 were disregarded as the care staff member mistakenly conducted the second-day assessment before the Aikomi session instead of after. In many cases, the participants had no recollection of the intervention the following day. Little inference can be drawn from the MENFIS data except that there appeared to be no adverse effects on the well-being of the people with dementia after using the Aikomi application.
Table 5. Well-being and behavior changes over the course of the intervention.

<table>
<thead>
<tr>
<th>Person</th>
<th>MENFIS score&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Care staff written comments (translation from Japanese)&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day 1</td>
<td>Day 2</td>
</tr>
<tr>
<td>1</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>28</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>10</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>22</td>
<td>26</td>
</tr>
<tr>
<td>13</td>
<td>27</td>
<td>12</td>
</tr>
<tr>
<td>14</td>
<td>24</td>
<td>27&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
<td>16&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>MENFIS: Mental Function Impairment Scale (6 noncognitive items). Day 1: evaluation the day before the session; day 2: evaluation on the day of the session after it was conducted; day 3: evaluation the day after the session.

<sup>b</sup>Translation by the research team.

<sup>c</sup>BPSD: behavioral and psychological symptoms of dementia.

<sup>d</sup>DLB: dementia with Lewy bodies.

<sup>e</sup>Recorded before the session was conducted.

Selected Case Studies

**Person 1**

Although not formally experiencing BPSD according to the NPI-NH, the care staff member said that person 1 usually displayed apathetic behavior, and the staff member was very surprised by the amount of self-initiated dialogue in response to the family and wartime navy service STIM, which was shown several times during the test. The STIM included a photograph of the ship he served on as well as a Japanese navy song. The person became conversational and could recall the names of former navy colleagues as well as events in detail, providing new information and anecdotes with every repetition of the STIM. The care staff used this information to continue to manually provide stimulation using photographs on the following day, which is likely to have contributed to the continuing improvement trend observed in the MENFIS score from the second to the third day. In contrast, person 2, who was his spouse and lived at the same facility but showed more advanced dementia symptoms, did not respond strongly to the same family STIM used for person 1. However, at the request of the care staff member, an additional session with the Aikomi application was conducted with both persons 1 and 2 watching the family STIM together. In this case, person 1 provided prompts to person 2, who showed a much stronger response than when she viewed the STIM with the care staff member.

**Person 10**

The caregiver reported that the person had shown agitated behavior in the period before the intervention but was able to come into the room and quickly settled when the session started. She responded to several of the STIMs and actively talked about several of the themes presented. As she was an excellent athlete during her school days, the program included a STIM video of a modern high school girls’ 100-meter race, in which she showed great interest, including making a humorous comment about the lack of clothes worn by current athletes. In addition, she showed quiet concentration on the STIM related to her hobbies of watercolor painting and hill walking and told the care staff member that she wanted to do both again and recalled the names of several mountains that she had climbed. During the smell intervention, she laughed and engaged with the care staff member and research team and appeared to enjoy the experience even though she did not recognize any of the smells. The care staff member said that her behavior during the intervention was not usual and that she was not agitated when she returned to her living area. However, this was not reflected in the MENFIS score.
scores, which were assessed by a different care staff member and showed no changes over the 3-day period.

**Person 13**

During the intervention, she showed good responses to STIMs about her hometown (Tokyo), her hobby of playing the koto (musical instrument), and a popular Japanese singer. She momentarily cried on 2 occasions when viewing her hometown and the Japanese singer, but her sadness did not persist and was followed by self-initiated talk to the care staff member about her wartime childhood and the Japanese singer’s unhappy life story. The care staff member reported that she usually could not show good concentration, and they were surprised that she was calm and could concentrate during the intervention and recall her memories. In the evening after the test, the care staff member reported that she said that she felt good and that other residents noticed that she was not as agitated as usual; however, on the day after the test, her agitated behavior returned to normal, and this pattern was reflected in the MENFIS scores.

**Discussion**

**Principal Findings**

Maintaining communication and engagement for persons with dementia is integral to caregiving, implementing person-centered care, and managing BPSD. It is also vital for facilitating the positive aspects of care that foster good QOL and sustainable caregiving relationships [45]. Over the last decade, several digital applications have become available to support communication and engagement for people with dementia [27], and to the authors’ knowledge, this pilot study is the first investigation of a digital application for personalized multi-sense stimulation conducted in Japan. All participants in this study were able to accept viewing the programs on the tablet, and these preliminary findings are broadly consistent with those of previous studies that demonstrate the importance of tailoring cognitive stimulation content to the individual profile and abilities of each person to obtain good engagement [28]. As reported for other applications, themes based on personal photographs or music often generated positive responses from participants, although for many people in the study, these themes did not generate the strongest responses during the intervention (Table 4). Instead, a wider range of themes specifically associated with the person’s lived experiences, such as wartime service (person 1), and interests such as gardening (person 5) and Western movies (person 9) were often the ones that resulted in prolonged and in-depth participant-initiated communication with care staff. This is in agreement with digital storytelling studies that have demonstrated the need to use a diverse range of relevant topics to adequately personalize interventions to obtain good engagement [29,30]. In addition, the benefits of using applications that combine both generic activity-promoting content and personally sourced content have been demonstrated with the CIRCA and CIRCUS application [46]. The role of the care staff was important to facilitate engagement in this study, especially to allow persons with dementia sufficient time to respond to the stimulation. It has been reported that caregivers sometimes do not allow sufficient time for the person with dementia to respond, which can lead to carer-directed engagement and suppression of the person with dementia’s own ability to initiate and maintain conversation [47]. In several cases, participants did not respond immediately to the stimulation content and required some time to view and “acclimatize” to the content before responding. This suggests that adapting the duration and complexity of the STIM according to the responses of each person may enhance their ability to respond and avoid premature transition to the next STIM, which will be explored in the future. The example of a husband and wife (persons 1 and 2) suggests the potential of people with dementia using the Aikomi application together without the active presence of care staff when the content has a high personal meaning for both. In total, 13% (2/15) of the participants, who had advanced dementia (persons 12 and 14), showed almost no engagement; however, good engagement was observed with another 13% (2/15) of the participants, who had low cognition (persons 6 and 10), suggesting that other factors in addition to cognition, such as content relevance, mood, and care staff behavior, may also be important for engagement.

One of the features of the Aikomi platform is that it allows for flexible navigation through a sequence of personalized stimulation themes that provide options for caregivers to select and adapt the content of the stimulation program according to the mood and responses of the user. The usefulness of this function was illustrated in person 1, who showed an unexpectedly strong reaction to the war-themed pictures triggering extended dialogue with the caregiver that was facilitated by several repetitions of the same STIM. In contrast, when person 13 started to make negative comments while viewing a STIM, it was possible to quickly change to a new STIM about her hobby, after which she started to make positive comments. Furthermore, the ability of the Aikomi application to easily modify and add new content enables the convenient creation of new personalized stimulation programs each time the Aikomi device is used. This is important for longer-term use to minimize or avoid the repeated use of the same content, which may lead to reduced interest from users and caregivers and has been found to be an issue for applications restricted to a fixed pool of content [33]. In addition, the flexible use of personally targeted activity content has been reported to promote curiosity and encourage self-directed learning in people with dementia [48] and suggests that personalizable digital cognitive stimulation applications such as Aikomi may be able to expand the range and depth of self-expression of people with dementia.

A secondary objective of this study was to explore the potential benefits of using paired smell and audiovisual stimulation to promote improved engagement compared with audiovisual stimulation alone. Currently, the use of smell stimulation in dementia care is limited and mainly focused on aromatherapy approaches to improve mood and reduce responsive behaviors using natural oils. However, other studies have shown that smell may play a role in triggering autobiographical and implicit memory [49], and it was thought that synthetic smells associated with daily life experiences could encourage not only the person with dementia but also the care staff member to share their own experiences, as well as being enjoyable. The engagement during paired smell and digital stimulation appeared to involve more smiling and laughter for both the participant and care staff.
member compared with responses to the audiovisual program alone. Interestingly, this was the case even when the person with dementia did not identify the smell and suggests that the combined use of digital audiovisual stimuli with smell may be more effective than smell alone [50].

From its inception, the Aikomi platform has aimed to develop AI capabilities to minimize the time and expertise required by non-specialist care staff and families to use personalized digital applications. This is a critical issue for the adoption and sustained use in dementia care, where caregivers are often older and not “digital natives” and have limited time or support to learn how to use and personalize applications [51]. Currently, most digital applications support personalization using two types of approaches: (1) supported collaboration with families to create bespoke interventions or (2) provision of predesigned content for on-demand selection during application use [28]. The bespoke approach often focuses on personal reminiscence and identity-reinforcing applications such as digital storytelling [29,30] and memory books [52,53], but preparation requires extensive family involvement over weeks or months, which may be challenging to sustain. Conversely, on-demand selection approaches, usually providing content curated by expert research teams, allow for immediate use and scalability but may fall short of generating sufficient interest and maintaining long-term use and be more suitable for social interaction [54] and activity-based applications such as iCST [33], music [55], and games [56]. The data obtained by digital technologies open up new opportunities to use machine learning to develop automation that can overcome these personalization barriers as well as optimize and adapt interventions. However, the lack of available high-quality and personalized data sources for people with dementia has limited progress in AI development for dementia care [35], especially for applications to support communication and engagement. To address this, Aikomi’s modular architecture facilitates the seamless capture of user data across key function domains: personal profile, content tagging, sequence ordering, and response analysis. This approach can combine the precision of the bespoke approach with the convenience of using the prepared on-demand content. In this pilot study, data from the interviews and the provided content were used with the prepared generic content to create bespoke programs in approximately 2 weeks, a reduction from the 6 to 8 weeks reported for digital storytelling [28]. In addition, it was possible to evaluate the performance of personalization using high-context data such as personal profiles, content and sequence attributes, and behavioral responses captured by the Aikomi platform. A related approach was reported by another personalized digital application called Scrapbook, which demonstrates the importance of obtaining multiple personal context–related data inputs to enable analysis [57]. Although AI system development was not the focus of this pilot study and personalization was conducted manually by the research team, this study demonstrates the potential of using the Aikomi platform as a tool to generate personalized data for AI development, and a preliminary investigation was conducted to create machine learning models to automate the personalization process, which is reported elsewhere [58]. In addition, chatbot technology, pioneered by a reminiscence intervention called ReminX [59], demonstrates an alternative trajectory for AI-driven personalization in dementia care.

Although no effects on BPSD were expected from this single-intervention study, the transient behavior changes reported by care staff for some persons (persons 1, 10, and 13) after using the Aikomi application suggest that there may be potential for investigating Aikomi to affect more lasting behavior changes related to BPSD. Although there is currently only a weak clinical evidence base to support the use of cognitive stimulation to manage BPSD, the use of personalized digital interventions embedded with data capture functions may offer the potential to not only create more effective therapeutics but also generate personalized monitoring data that can provide more robust clinical evidence. In the last few years, digital therapeutics (DTX) has emerged as a new category of regulatory-approved medical products that are distinct from drugs and medical devices [60,61]. To date, no DTX interventions for dementia have received regulatory approval, but several applications have been developed for supportive care [62]. DTX for dementia was pioneered by ReminX [59], which was designated as a breakthrough medical device by the Food and Drug Administration (FDA), and more recently by CST Assistant [63], which is a clinical evidence–supported CST-based game application that is now commercially available in Europe. Given the ongoing challenges in developing effective and affordable drug therapies for dementia and BPSD, digital interventions for personalized cognitive stimulation such as Aikomi may have potential for clinical development as DTX and offer nondrug options for the management of BPSD and improvement in QOL. In addition to the primary objective of addressing therapeutic goals, the importance of preserving care relationships and creating opportunities for positive aspects of care for caregivers is also gaining increasing attention in dementia care [64,65]. With the growing shortage of professional caregivers, which is particularly acute in Japan, digital technologies that can promote meaningful engagement and improve QOL for both people with dementia and caregivers may become important tools to foster greater participation by family and informal caregivers in caring for their loved ones.

Limitations

As this study was the first evaluation of the Aikomi application conducted in a care setting and limited to a single intervention to confirm its safety and acceptability for people with dementia, all inferences are preliminary and need to be confirmed via further multiple-intervention studies. Furthermore, although the aim was for Aikomi to be used independently by the care staff, the research team was required to be present during the interventions for this pilot study, which was a potential source of bias. This was necessary as the Aikomi platform was still a prototype, and the care staff were not familiar with operating digital technologies and required support to use Aikomi and overcome any technical difficulties. In fact, no application-related technical problems were encountered during the intervention, and connectivity issues were largely avoided by using tablets with SIM cards and addressed before the intervention. However, the provision of appropriate training and technical support for care staff to set up, use, and maintain the Aikomi application was not investigated in this pilot study.
Another limitation was that most staff members were unfamiliar with conducting EPWDS and MENFIS evaluations, and the evaluations of some participants were conducted by multiple care staff members because of shift changes, which may have led to some inconsistencies in the results. Furthermore, the EPWDS is not yet available in Japanese, and an unvalidated translation prepared by the research team was used. These limitations need to be addressed in future studies.

Future Research

The next step is to conduct multiple-intervention studies to investigate the effects of longer-term use of the Aikomi application when the program content of each intervention is adapted based on the behavioral responses of the person with dementia. The data derived from these studies will become the basis for developing machine learning models to create algorithms that can optimize personalization and increase the convenience for the care staff to use the system. To generate evidence to support the use of the Aikomi application in dementia care, it is necessary to conduct clinical trials to evaluate its effects on QOL and BPSD, from which its potential for further development as a digital therapeutics can be assessed. Separately, use by family caregivers in their own homes will be investigated to obtain feedback and data that will guide the development of the Aikomi application for community use. In addition, further work is needed to increase the pool of diverse and culturally relevant content that has been curated for people with dementia to reduce the lead time required to prepare stimulation programs. In this pilot study, all non–family-derived content was selected or created by the research team from publicly available sources, and copyright issues were not considered because of the noncommercial nature of this research. However, ensuring copyright compliance for the use of all digital media content, for both content owners and content providers, is a significant issue that must be addressed before the commercial deployment of personalized cognitive stimulation approaches. Finally, to enable more convenient use of smell stimuli with cognitive stimulation, automated smell delivery devices such as diffusers should be investigated for integration with the Aikomi platform.

Conclusions

This pilot study demonstrated that the Aikomi application was able to create personalized cognitive stimulation programs that were acceptable for use in Japanese care homes and may have the potential to promote communication between people with dementia and their care staff. The use of smell stimuli paired with audiovisual stimulation was found to promote enjoyable interactions for many users. In addition, the Aikomi platform captured several types of personalized data, including the behavioral responses during the intervention, which enabled a detailed analysis of the stimulation content preferences of each person with dementia. These results indicate that the Aikomi application has the potential to be used as a tool to provide personalized cognitive stimulation and also generate high-context data suitable for the future development of AI systems to automate the personalization process. Further research will be conducted to develop the Aikomi application as a communication tool that can be easily used by nonspecialist care staff and families in residential and community care settings to enhance care relationships and positive aspects of care and aid the therapeutic management of BPSD.

Acknowledgments

The authors would like to thank all the people with dementia and their families and participating care staff for their support and cooperation in participating in this study. The help and advice of the care managers at the participating care homes is gratefully appreciated. Ken Inoue (chief executive officer, Promotool Corporation) is thanked for his kind advice and supply of the smell sticks.

Data Availability

The data sets generated and analyzed in this study are not publicly available as the data contain information that could compromise the privacy of the research participants. The data are available from the corresponding author upon reasonable request.

Authors’ Contributions

NH, SG, and SKP collaboratively developed the Aikomi technology platform. KM and TO provided guidance throughout the development as well as advice and supervision of the pilot study in Japanese care homes. NH supported the conduct of the pilot study.

Conflicts of Interest

NH is a cofounder and chief scientific officer of Aikomi. NH and the SBX Corporation are shareholders of Aikomi.

Multimedia Appendix 1

Full details of engagement during the session with the Aikomi device.

References
33. Rai HK, Schneider J, Orrell M. An individual cognitive stimulation therapy app for people with dementia: development and usability study of thinkability. JMR Aging 2020 Nov 16;3(2):e17105 [FREE Full text] [doi: 10.2196/17105] [Medline: 33196451]
34. Rai HK, Schneider J, Orrell M. An individual cognitive stimulation therapy app for people with dementia and their carers: protocol for a feasibility randomized controlled trial. JMR Res Protoc 2021 Apr 08;10(4):e24628 [FREE Full text] [doi: 10.2196/24628] [Medline: 33830058]
37. Promotool. URL: https://www.promotool.jp [accessed 2023-12-19]
Abbreviations

AI: artificial intelligence  
BPSD: behavioral and psychological symptoms of dementia  
CMS: content management system  
CST: cognitive stimulation therapy  
DTX: digital therapeutics  
EPWDS: Engagement of a Person with Dementia Scale

https://formative.jmir.org/2024/1/e51732
FDA: Food and Drug Administration
HDS-R: revised Hasegawa Dementia Scale
iCST: individual cognitive stimulation therapy
MENFIS: Mental Function Impairment Scale
MMSE: Mini-Mental State Examination
NM: Nishimura Mental State Scale for the Elderly
NPI-NH: Neuropsychiatric Inventory–Nursing Home version
QOL: quality of life

©Nick Hird, Tohmi Osaki, Samik Ghosh, Sucheendra K Palaniappan, Kiyoshi Maeda. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Designing and Developing a Mobile App for Management and Treatment of Gestational Diabetes in Nepal: User-Centered Design Study

Aarthi Shanmugavel¹, MPH; Prabin Raj Shakya², MS; Archana Shrestha³,⁴,⁵, PhD; Jyoti Nepal⁴, MPH; Abha Shrestha⁶, MD; Jean-Francois Daneault⁷, PhD; Shristi Rawal⁸, PhD

¹Department of Health Informatics, School of Health Professions, Rutgers, The State University of New Jersey, Piscataway, NJ, United States
²Biomedical Knowledge Engineering Lab, Department of Dentistry, Seoul National University, Seoul, Democratic People's Republic of Korea
³Institute for Implementation Science and Health, Kathmandu, Nepal
⁴Department of Public Health, Kathmandu University School of Medical Sciences, Dhulikhel, Nepal
⁵Department of Chronic Disease and Epidemiology, Center of Methods for Implementation and Prevention Science, Yale School of Public Health, New Haven, CT, United States
⁶Department of Obstetrics and Gynecology, Dhulikhel Hospital, Dhulikhel, Nepal
⁷Department of Rehabilitation and Movement Sciences, School of Health Professions, Rutgers University, Newark, NJ, United States
⁸Department of Clinical and Preventive Nutrition Sciences, School of Health Professions, Rutgers, The State University of New Jersey, Newark, NJ, United States

Corresponding Author:
Shristi Rawal, PhD
Department of Clinical and Preventive Nutrition Sciences
School of Health Professions
Rutgers, The State University of New Jersey
Stanley S Bergen, Jr Building, Suite 923
65 Bergen Street
Newark, NJ, 07107
United States
Phone: 1 973 972 2710
Fax: 1 973 972 7403
Email: shristi.rawal@rutgers.edu

Abstract

Background: Mobile apps can aid with the management of gestational diabetes mellitus (GDM) by providing patient education, reinforcing regular blood glucose monitoring and diet/lifestyle modification, and facilitating clinical and social support.

Objective: This study aimed to describe our process of designing and developing a culturally tailored app, Garbhakalin Diabetes athawa Madhumeha—Dhulikhel Hospital (GDM-DH), to support GDM management among Nepalese patients by applying a user-centered design approach.

Methods: A multidisciplinary team of experts, as well as health care providers and patients in Dhulikhel Hospital (Dhulikhel, Nepal), contributed to the development of the GDM-DH app. After finalizing the app’s content and features, we created the app’s wireframe, which illustrated the app’s proposed interface, navigation sequences, and features and function. Feedback was solicited on the wireframe via key informant interviews with health care providers (n=5) and a focus group and in-depth interviews with patients with GDM (n=12). Incorporating their input, we built a minimum viable product, which was then user-tested with 18 patients with GDM and further refined to obtain the final version of the GDM-DH app.

Results: Participants in the focus group and interviews unanimously concurred on the utility and relevance of the proposed mobile app for patients with GDM, offering additional insight into essential modifications and additions to the app’s features and content (eg, inclusion of example meal plans and exercise videos). The mean age of patients in the usability testing (n=18) was 28.8 (SD 3.3) years, with a mean gestational age of 27.2 (SD 3.0) weeks. The mean usability score across the 10 tasks was 3.50 (SD 0.55; maximum score=5 for “very easy”); task completion rates ranged from 55.6% (n=10) to 94.4% (n=17). Findings from the usability testing were reviewed to further optimize the GDM-DH app (eg, improving data visualization). Consistent with social cognitive theory, the final version of the GDM-DH app supports GDM self-management by providing health education.
and allowing patients to record and self-monitor blood glucose, blood pressure, carbohydrate intake, physical activity, and gestational weight gain. The app uses innovative features to minimize the self-monitoring burden, as well as automatic feedback and data visualization. The app also includes a social network “follow” feature to add friends and family and give them permission to view logged data and a progress summary. Health care providers can use the web-based admin portal of the GDM-DH app to enter/ review glucose levels and other clinical measures, track patient progress, and guide treatment and counseling accordingly.

**Conclusions:** To the best of our knowledge, this is the first mobile health platform for GDM developed for a low-income country and the first one containing a social support feature. A pilot clinical trial is currently underway to explore the clinical utility of the GDM-DH app.

**KEYWORDS**
mHealth; mobile health; gestational diabetes; telehealth; usability testing; LMICs; low- and middle-income countries; user-centric design; social cognitive theory; South Asians; maternal health; diabetes; diabetes mellitus; daily glucose monitoring; hospital; medical institution; health center; clinical utility; Nepal; low income; clinical trial; focus group; interview; health care provider; medical practitioner; mobile app; application; digital health; app; apps; health education; web based; self-monitoring; glucose; physical activity; intervention

**Introduction**

Gestational diabetes mellitus (GDM), defined as hyperglycemia with onset during pregnancy, is a major public health issue worldwide. South Asians, who represent approximately one-fourth of the world’s population, are at a disproportionately higher risk of GDM [1-3], and the prevalence of GDM is increasing rapidly in South Asian countries, including Nepal [3-5]. Geographically situated between the two epicenters of the global diabetes epidemic, India and China [6], Nepal has a reported GDM prevalence ranging from 6.6% to 28% [7-9]. These estimates are alarming as GDM is associated with serious adverse perinatal outcomes and unfavorable long-term cardiometabolic consequences in both women and their children [10-14]. Although the short-term health and economic burden of GDM is substantial [15,16], its long-term implications are even more concerning, particularly among South Asian populations who are known to develop cardiometabolic complications at a relatively lower BMI than people with European ancestry [17-19]. Among women with GDM, those with South Asian ethnicity are also known to have a significantly higher risk of developing type 2 diabetes (T2D) compared to other ethnic groups [3,20]. Scalable and cost-effective solutions are thus needed to address the growing burden of GDM and its sequelae, particularly in low-resource South Asian countries, such as Nepal.

Successful GDM management relies on patient adherence to a complex care regimen, including dietary modification, adequate physical activity, weekly-to-biweekly antenatal follow-ups, and regular blood glucose monitoring and logging. Providing dietary and physical activity recommendations is a critical part of GDM management, but in resource-limited settings, such as Nepal, time for diet/lifestyle counseling often competes with other components of care. Face-to-face counseling for diet/lifestyle is also fraught with low participation rates and high attrition, as it is resource intensive for the health care providers and poses time and travel barriers for the patients [21,22]. Additionally, in many countries like Nepal, GDM counseling is only performed once, which is not conducive to facilitating a meaningful behavior change in diet and lifestyle. Mobile health (mHealth) technology provides new opportunities to circumvent these challenges [23] and support the treatment and management of GDM in low-resource settings. Mobile apps can aid in the management of GDM by providing patient education, reinforcing regular glucose monitoring and diet/lifestyle modification, and allowing health care providers to communicate and exchange health information with patients [24]. Mobile technology may offer cost-effective strategies to improve outcomes in patients with GDM by augmenting clinical care and empowering patients with GDM to self-manage their condition, yet this approach has not been tested previously in any low-income country [24].

App-based lifestyle interventions for GDM management are not common, especially in low-income countries, such as Nepal, where its prevalence is rapidly increasing [23,25,26]. To address this gap, our goal was to develop a mobile app that supports self-management and treatment among women with GDM in Nepal. In addition to making the app culturally tailored, our priority was also to design an app that matches the user needs and technological sophistication of the target users. Thus, taking a user-centered design approach [27], we developed the Garbhakalin Diabetes athawa Madhumeha—Dhulikhel Hospital (GDM-DH) app in collaboration with our target users, patients with GDM and providers, in Dhulikhel Hospital, a flagship university hospital in Dhulikhel, Nepal. Consistent with Bandura’s social cognitive theory (SCT) framework [28], the GDM-DH app supports GDM management by providing patient education, reinforcing regular blood glucose/carbohydrate monitoring, increasing self-efficacy for diet/lifestyle modification, and facilitating clinical and social support. Here, we aim to describe the app development process and features of the GDM-DH app.

**Methods**

**Overall Study Design**

The study was conducted at Dhulikhel Hospital, a community-based tertiary-level university hospital of Kathmandu University (Nepal). We took a user-centered design approach to develop a culturally tailored mobile app (GDM-DH).
for management of patients with GDM at the hospital. Figure 1 outlines the steps in GDM-DH app development. In the qualitative/requirement-gathering phase, patients with GDM were recruited for a focus group and structured interviews to show them the app prototype and obtain their feedback on its features and functions. Key informant interviews (KIIs) were conducted with clinicians and patients’ spouses. Incorporating and revising the app prototype based on user input, we built a minimum viable product (MVP), after which additional patients with GDM were recruited for usability testing including the think-aloud protocol [29]. The final GDM-DH app was developed following an iterative process of product design and user testing.

Figure 1. Schematic representation of the user-centered approach for GDM-DH app development among target users (women with gestational diabetes) in Dhulikhel Hospital, Nepal. GDM: gestational diabetes mellitus; GDM-DH: Garbhakalin Diabetes athawa Madhumeha—Dhulikhel Hospital; KII: key informant interview.

Recruitment Procedure and Inclusion Criteria
Located in a periurban setting, about 20 km from the capital city of Kathmandu in Dhulikhel, Nepal, Dhulikhel Hospital has a catchment population of 1.9 million people and delivers approximately 3000 babies annually. All pregnant women receiving antenatal care at the Obstetric Outpatient Department at Dhulikhel Hospital undergo routine screening for GDM at 24-28 weeks of gestation. Inclusion criteria were pregnant women who (1) received antenatal care at Dhulikhel Hospital, (2) received a GDM diagnosis (within the preceding year), (3) owned a smart phone, and (4) could understand and read Nepali. Patients with a confirmed GDM diagnosis were recruited into the study with the help of a senior obstetrician-gynecologist (OB-GYN; coinvestigator in the study) and other staff in the OB-GYN department at Dhulikhel Hospital. A convenience sampling strategy was used to recruit participants meeting the aforementioned inclusion criteria for usability testing (n=18) and qualitative user research (n=19); for the latter, participants were recruited until data saturation was achieved.

Ethical Considerations
The study protocol was approved by the Rutgers Newark Health Sciences Institutional Review Board (Pro2019001883) and the Ethical Review Board of the Nepal Research Health Council (NHRC; registration number 735/2019). Signed written informed consent was obtained from all participants by the research assistant at Dhulikhel Hospital. To ensure participant confidentiality, all documents including participant identifiers, such as the master list and consent forms, are stored separately in a locked cabinet and in a secure password-controlled Health Insurance Portability and Accountability Act (HIPAA)-compliant BOX folder. Only select research staff have access to the documents and folders containing participant identifiers and data. The participants (women with GDM and their spouses) received a mobile recharge card worth Nepalese rupees (NRs) 500 (US $3.77) to compensate for their time for the interview/focus group or usability testing.

App Development Stages

Prototype Development
A multidisciplinary team including experts in GDM, mHealth, and behavior and implementation sciences, as well as health care providers and patients at Dhulikhel Hospital, contributed to the development of the GDM-DH app. Content modules and features to be included in the app prototype were selected based on a literature review, theory-based behavioral strategies, discussions with subject matter experts, and international recommendations and guidelines (including the Package of Essential Noncommunicable [PEN] disease interventions for primary health care in low-resource settings) [30]. A series of meetings and a full-day workshop were conducted with the research team to select and finalize the features of the app prototype. During the meetings, app features were selected based on expected user needs, alignment with theory-based constructs for behavior change, and the logistical and economic feasibility of incorporating these features in the app. The meetings and deliberations spanned over several weeks until differences were resolved and consensus was reached.

Qualitative User Research
After finalizing the app’s content and features, a focus group and structured interviews were conducted to explore the perceived barriers to and facilitators of GDM management and to seek feedback on the GDM app prototype. A total of 12 women with a GDM diagnosis (either current or in the preceding 1 year) were recruited from Dhulikhel Hospital, 4 (33.3%) of whom participated in a focus group and the remaining 8 (66.7%) in structured interviews. KIIs were also conducted with health care providers (n=5) and spouses of patients with GDM (n=2). All interviews were audio-taped and transcribed verbatim. A
and researchers. The features and functionalities of the GDM-DH app were guided by Bandura’s SCT [28], which was selected as it has been widely applied in the dietary/lifestyle management of chronic health conditions [35] and is shown to be a suitable framework for promoting healthy behaviors among pregnant women, including those with GDM [36,37]. We focused our intervention modules on the SCT constructs of self-efficacy (confidence in one’s ability to take action and overcome perceived barriers to a behavior change), self-regulation/self-control (ability to understand and manage feelings, behaviors, and actions to achieve goals), behavioral capabilities (knowledge and skills needed to perform a given behavior), reinforcements (responses to a person’s behavior that increase or decrease the likelihood of occurrences), and outcome expectations/expectancies (anticipated outcomes of a behavior and values a person places on the probable outcomes of a behavior) [38]. Behavior change techniques (BCTs) [39] for the GDM-DH intervention content were selected based on the published literature [40] that maps the BCTs with the SCT constructs for behavior change (eg, the SCT of information about health consequences aligns with the SCT construct of outcome expectations). The SCT constructs observational learning (acquiring a new behavior by watching someone else performing it and observing their outcomes) and environment (physically external factors that can influence a behavior) were not targeted, as it was not feasible to achieve them at this time using the mobile app.

Using the SCT framework for behavior change, we decided that the content and features included in the GDM-DH app would support self-management of GDM by (1) providing health education, (2) helping patients identify and set target health goals (for diet, physical activity, and glucose levels), (3) enhancing their self-efficacy to meet target goals, and (4) facilitating desired support from family members. In SCT, self-monitoring of behavior is the first and most important step in self-regulating appropriate behavior changes [28]. Self-monitoring is also known to be a powerful behavior change strategy for changing diet and physical activity [41]. Hence, we decided that the core features of the GDM-DH app would facilitate the users to record and self-monitor their blood glucose, blood pressure, carbohydrate intake, physical activity, and gestational weight gain (GWG). The app would use automatic feedback and data visualization to aid in self-monitoring, as well as innovative technological features to minimize the self-monitoring burden, including visual aids for estimating carbohydrate portion sizes and smartphone GPS and accelerometer sensors for obtaining physical activity data [42]. Health care providers would be able to use the web-based portal of the GDM-DH app to enter/review blood glucose readings, track patient progress, and accordingly guide treatment and counseling [42].

In many South Asian countries, women are not the sole health decision makers, with mothers-in-law and husbands having a strong influence on their health decisions during pregnancy [43-45]. Additionally, family members are closely involved in a pregnant woman’s food selection and preparation, thus influencing her dietary behaviors. Hence, the team decided that the GDM-DH app would be designed to garner social support...
All participants agreed that the proposed mobile app and features would be useful and relevant to women with GDM. They believed it would help overcome existing barriers by empowering pregnant women with information and tools to manage GDM and track their progress.

If we had seen this app before, I think we would have been able to control blood sugar levels, and we would have been able to plan. I think after seeing this, it would have helped, it would be useful. [Spouse of a patient with GDM on the GDM-DH mobile app]

Just knowing that this app on its own records and tells you about your physical activity makes us alert...we will know how much more activity we need to do...it makes it easier. [Patient with GDM on the GDM-DH mobile app]

However, both patients with GDM and health professionals requested more content with respect to medical management and diet/lifestyle modification for GDM. Based on findings from the qualitative study, we changed some app features and design elements (e.g., data visualization), in addition to modifying the educational materials and other resources to further tailor the GDM-DH app culturally. For example, the educational modules were revised to address specific cultural and social challenges faced by our patients (e.g., food-centered festivals, long-held dietary/cultural practices surrounding pregnancy), and appropriate strategies were provided to problem-solve around these barriers. Based on our target users’ suggestions, we added example meal plans with locally available and culturally staple foods, video demonstrations of safe and culturally relevant exercises during pregnancy (e.g., yoga, mild hiking, walking), and revised visual aids for carbohydrate estimation to include standardized pictures of staple Nepalese foods with common portion sizes shown in locally used utensils, such as plates, bowls, and cups.

**Usability Testing**

In total, 18 newly diagnosed patients with GDM participated in the usability testing [29] with the MVP. The mean age of patients in the usability testing was 28.8 (SD 3.3) years. All patients were married, and slightly more than half (n=10, 55.6%) were homemakers. The mean gestational age was 27.2 (SD 3.0) weeks, and the average number of years of schooling was 13.3 (SD 2.8) years.

Results from the think-aloud protocol are described in Table 1. The mean usability score across the 10 tasks was 3.50 (SD 0.55; maximum score=5 for very easy). The task completion rates ranged from 55.6% (n=10) to 94.4% (n=17) across the 10 tasks, with the lowest completion rate for the task requiring the patients to look up their next scheduled appointment on the app. All patients except 1 (5.6%) were able to successfully complete tasks requiring them to enter their weight and systolic and diastolic pressure into the app.
Table 1. Usability testing of key important features of the GDM-DH\(^a\) app among target users (women with GDM\(^b\)) in Dhulikhel Hospital, Nepal (n=18).

<table>
<thead>
<tr>
<th>Usability testing task</th>
<th>Successful completion, n (%)</th>
<th>Very difficult, n (%)</th>
<th>Difficult, n (%)</th>
<th>Normal, n (%)</th>
<th>Easy, n (%)</th>
<th>Very easy, n (%)</th>
<th>Score(^c), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enter fasting blood glucose levels.</td>
<td>16 (88.9)</td>
<td>2 (11.1)</td>
<td>0</td>
<td>5 (27.8)</td>
<td>11 (61.1)</td>
<td>2 (11.1)</td>
<td>3.8 (0.6)</td>
</tr>
<tr>
<td>Enter postprandial blood glucose levels and view the glucose chart.</td>
<td>16 (88.9)</td>
<td>2 (11.1)</td>
<td>0</td>
<td>5 (27.8)</td>
<td>13 (71.2)</td>
<td>0</td>
<td>3.7 (0.5)</td>
</tr>
<tr>
<td>Enter the systolic blood pressure level.</td>
<td>17 (94.4)</td>
<td>1 (5.6)</td>
<td>0</td>
<td>6 (33.3)</td>
<td>10 (55.6)</td>
<td>2 (11.1)</td>
<td>3.8 (0.6)</td>
</tr>
<tr>
<td>Enter the diastolic blood pressure level and view the blood pressure chart.</td>
<td>17 (94.4)</td>
<td>1 (5.6)</td>
<td>0</td>
<td>5 (27.8)</td>
<td>13 (71.2)</td>
<td>0</td>
<td>3.7 (0.5)</td>
</tr>
<tr>
<td>Open a video on GDM nutrition.</td>
<td>16 (88.9)</td>
<td>2 (11.1)</td>
<td>0</td>
<td>3 (16.7)</td>
<td>12 (66.7)</td>
<td>1 (5.6)</td>
<td>3.7 (0.8)</td>
</tr>
<tr>
<td>Add a friend or family member in the app.</td>
<td>12 (66.7)</td>
<td>6 (33.3)</td>
<td>1 (5.6)</td>
<td>4 (22.2)</td>
<td>5 (27.8)</td>
<td>7 (38.9)</td>
<td>1 (5.6)</td>
</tr>
<tr>
<td>Enter weight.</td>
<td>17 (94.4)</td>
<td>1 (5.6)</td>
<td>0</td>
<td>4 (22.2)</td>
<td>12 (66.7)</td>
<td>2 (11.1)</td>
<td>3.9 (0.6)</td>
</tr>
<tr>
<td>Find out the daily step count from today.</td>
<td>12 (66.7)</td>
<td>6 (33.3)</td>
<td>1 (5.6)</td>
<td>3 (16.7)</td>
<td>6 (33.3)</td>
<td>7 (38.9)</td>
<td>1 (5.6)</td>
</tr>
<tr>
<td>Figure out how many carbohydrates were consumed at breakfast today.</td>
<td>12 (66.7)</td>
<td>6 (33.3)</td>
<td>0</td>
<td>4 (22.2)</td>
<td>6 (33.3)</td>
<td>8 (44.4)</td>
<td>0</td>
</tr>
<tr>
<td>Find out when the next appointment is.</td>
<td>10 (55.6)</td>
<td>8 (44.4)</td>
<td>2 (11.1)</td>
<td>5 (27.8)</td>
<td>5 (27.8)</td>
<td>6 (33.3)</td>
<td>2.8 (1.0)</td>
</tr>
</tbody>
</table>

\(^a\)GDM-DH: Garbhakalin Diabetes athawa Madhumeha—Dhulikhel Hospital.

\(^b\)GDM: gestational diabetes mellitus.

\(^c\)Patients were asked to rate the difficulty of completing each task on a 5-point scale ranging from 5 for “very easy” to 1 for “very difficult.” The mean (SD) score represents an average score for the corresponding task across all patients.

**Modifications to the GDM-DH App After Usability Testing**

Findings from the focus group, interviews, and usability testing were reviewed to identify recurring themes of feedback with respect to the GDM-DH app’s content, usability, navigation, and functionalities. These findings allowed us to gain insight into participants’ thought processes with regard to app use, places where they encounter difficulties, and ways to improve the app’s usability. As shown in Textbox 1, these themes served as valuable insights that were used by the app developer to address the key usability barriers and participants’ preferences/feedback and develop a final version of the GDM-DH app for testing in a pilot clinical trial. For example, considering that nearly half of the usability testing participants struggled in identifying their upcoming antenatal appointments, we decided that the upcoming appointments would be shown in a list, in addition to the calendar. Additionally, we included a comprehensive video tutorial on how to navigate the GDM-DH app and use its features. In the web-based portal, in addition to the features decided on by the research team and software development company, we incorporated new features requested by the providers, including a patient finder tool, data export customization, and additional analytics in the dashboard, such as the average number of antenatal visits per patient and the percentage of patients under medical therapy for GDM.
Textbox 1. Modifications requested and incorporated in the Garbhakalin Diabetes athawa Madhumeha—Dhulikhel Hospital (GDM-DH) app based on user research and usability testing among target users (women with gestational diabetes mellitus [GDM]) in Dhulikhel Hospital, Nepal.

**Educational resources**
- Use simple and clear language in educational resources, avoiding jargon and using terms understandable to the target users.
- Provide clear guidance on healthy food selections and appropriate portion sizes, including example meal plans with locally available foods.
- Clarify common misconceptions around diet and physical activity during pregnancy.
- Address specific cultural and social challenges faced by target users with respect to diet/lifestyle modification and provide appropriate strategies to problem-solve around those barriers.
- Add information about the signs and symptoms of hypoglycemia, along with practical strategies for effective management.
- Clarify criteria for when insulin or medication is indicated for GDM.
- Add examples of physical activity and exercise videos that are appropriate for pregnant women.
- Include information about contraindications for physical activity during pregnancy, including warning signs to stop exercising.
- Add an educational module on how family members can provide support to women with GDM.

**App interface and features**
- Use bigger font sizes; enlarge and make the images clearer.
- Revise the visual aids to include standardized pictures of common food portion sizes in locally used utensils, such as plates, bowls, and cups.
- Add reminders to input blood glucose, weight, diet, and blood pressure data.
- Use pop-ups to confirm data input and avoid double entries.
- Use bar graphs instead of line graphs to present blood glucose, weight, and blood pressure visualizations.
- Make glucose, weight, diet, and blood pressure data visible both as a list and in graphical format.
- Show upcoming appointments in a list, in addition to the calendar.
- Add the hospital hotline number under Help and Support.
- Include a video tutorial on how to navigate the GDM-DH app and use its features.

**Web-based portal**
- Include a patient finder feature to search for a patient quickly.
- Add a function to customize data export based on specific parameters and layout.
- Add dashboard analytics based on patient data (eg, average number of antenatal visits per patient, percentage of patients under medical therapy for GDM)

The decision to retain or disregard the requested modifications was based on their potential influence on behavioral and clinical outcomes, budget feasibility, the implementation time frame, and their potential impact on the app’s scalability in the future. For example, meal plans were added because they could be easily incorporated with minimal time and cost but would have maximal health gains. However, our target users also suggested that we add a feature to connect and chat with other app users with GDM, as well as a platform to communicate directly with health care providers via the app; these features were not added due to logistical, technical, and funding constraints but may be considered in the future.

**Final GDM-DH App**
Based on the SCT framework, the final version of our culturally tailored GDM-DH app supports GDM management by providing patient education, reinforcing regular blood glucose/carbohydrate/weight monitoring, increasing self-efficacy for diet/lifestyle modification, and facilitating clinical and social support. A detailed description and justification of the app features as well as their alignment with the SCT constructs are provided in Table 2.
Table 2. Description and justification of the GDM-DH\(^a\) app features designed for self-management and treatment of Nepalese women with GDM\(^b\).

<table>
<thead>
<tr>
<th>Feature and description</th>
<th>Rationale</th>
<th>SCT(^c) constructs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Educational modules</strong></td>
<td>The level of knowledge about GDM is significantly associated with self-management efficacy and glycemic control [47–49]. It also facilitates better information retention as patients can go through the lessons at their own pace and revisit them at their convenience.</td>
<td>• Self-efficacy  • Behavior capabilities  • Outcome expectancies</td>
</tr>
<tr>
<td>Educational modules consist of text- and image-based materials and brief videos covering various health and nutrition topics related to GDM and its management.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Blood glucose monitoring</strong></td>
<td>Self-monitoring of glucose levels is associated with an increase in self-efficacy and better glycemic control [50]. Data visualizations increases patient awareness and helps health care providers with timely and informed clinical decision-making.</td>
<td>• Self-regulation  • Self-efficacy</td>
</tr>
<tr>
<td>Blood glucose levels can be logged in for fasting and postprandial levels 3 times daily. These data are displayed in color-coded graphs (red and green bars for above- and within-target ranges, respectively).</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Carbohydrate monitoring</strong></td>
<td>It builds the user’s self-efficacy for understanding and changing their carbohydrate intake patterns.</td>
<td>• Self-regulation  • Behavior capabilities</td>
</tr>
<tr>
<td>The app incorporates standardized pictures of local Nepalese foods with common portion sizes to help the user estimate and track calories and carbohydrate (grams) in their meals.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Blood pressure monitoring</strong></td>
<td>Data visualization increases patient awareness and helps health care providers with timely and informed clinical decision-making.</td>
<td>• Self-regulation  • Self-efficacy</td>
</tr>
<tr>
<td>Users can log/track their systolic and diastolic blood pressure. These data are displayed in color-coded graphs (red and green bars for above- and within-target ranges, respectively).</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GWG(^d)</strong></td>
<td>Weight monitoring builds the user’s self-efficacy for understanding and managing their GWG.</td>
<td>• Self-regulation  • Self-efficacy</td>
</tr>
<tr>
<td>Based on weights entered by the user, the app creates a graph comparing the user’s weekly GWG rate with the recommended guidelines for optimal GWG, depending on the pre-pregnancy BMI.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Physical activity</strong></td>
<td>It builds the user’s self-efficacy for understanding and changing their physical activity patterns.</td>
<td>• Self-regulation  • Self-efficacy</td>
</tr>
<tr>
<td>The app integrates with the Google-Fit app to pull and graph physical activity data, including the step count.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Appointment reminder</strong></td>
<td>The reminder system enables patient adherence to the antenatal care regimen [51].</td>
<td>• Reinforcement</td>
</tr>
<tr>
<td>The app has an in-built calendar, which the users can use to record and view upcoming antenatal appointments.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Social network</strong></td>
<td>The app helps a user garner social support from friends/family and offers a source of accountability, motivation, and shared experience.</td>
<td>• Reinforcement</td>
</tr>
<tr>
<td>Via a social network “follow” feature, the patient is able to list 1 or more friends/family members as contacts in the app and give them the permission to view their logged data or progress summary.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Web-based portal</strong></td>
<td>It streamlines the providers’ workflow, as they can quickly look at patient data visualizations to understand patient behaviors and progress and accordingly guide their treatment and counseling.</td>
<td>• Reinforcement</td>
</tr>
<tr>
<td>Health care providers can use the web-based admin portal to register a new patient, as well as enter, update, or review clinical and other patient-related information (glucose/blood pressure/weight/diet, clinical history/notes).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)GDM-DH: Garbhakalin Diabetes athawa Madhumeha—Dhulikhel Hospital.

\(^b\)GDM: gestational diabetes mellitus.

\(^c\)SCT: social cognitive theory.

\(^d\)GWG: gestational weight gain.
Mobile App
The mobile app, which is patient facing, includes 6 feature icons on its home page: (1) Blood Glucose, (2) Food Intake, (3) Blood Pressure, (4) Weight, (5) Physical activity, and (6) Appointment (Figure 2). Using these features, the app allows patients to record and self-monitor their blood glucose, blood pressure, carbohydrate intake, physical activity, and GWG and track them over time. The app has a goal-setting feature and uses innovative technological features to minimize the self-monitoring burden, such as visual aids for carbohydrate estimation and integration with the Google-Fit app to automatically log physical activity data. Based on the data entered, the app provides automatic feedback about blood glucose, blood pressure, and GWG via a feedback engine that compares the user data to existing guidelines and recommendations. The app also generates visual displays summarizing their blood glucose, blood pressure, diet, physical activity, and weight patterns, allowing the user to easily monitor their alignment and progress toward target goals. In addition to the self-monitoring features, multimedia video- and text-based modules are included in the app as educational resources. The GDM-DH app also includes a social network “follow” feature, allowing the user to list 1 or more friends/family members as contacts in the app and give them the permission to view their logged data or progress summary. The app has an in-built calendar, which the users can use to record and view upcoming antenatal appointments.
Web-Based Portal

The web-based portal can be securely accessed by researchers and health care providers from any device that supports modern web browsers. The web-based portal has features for patient management, data capture and review, and data dashboard/visualization. Health care providers can register a new patient; enter, update, or review clinical and other patient-related information (patient vitals, measurements, clinical notes, medications, etc); and schedule or make changes in appointments. Using Apache Kafka, the web-based portal syncs with the mobile app and allows providers to access data and graphs summarizing the patient’s diet, physical activity, weight, blood pressure, and blood glucose patterns. This streamlines the providers’ workflow and allows them to easily track patient progress and accordingly guide their treatment and counseling.
Researchers and admin users can use the web-based portal to add new users, add/update the app modules/images/visualization, and audit changes made by users.

**Data Security**

Data from the mobile app are stored in a HIPAA-compliant, secure server hosted by Amazon Web Services. MongoDB is used to implement the database service, which is a free, open source, no–Structured Query Language (no-SQL) database program tailored to support big data. It is encrypted and access-controlled using tokens to ensure it cannot be accessed outside the app. Apache Kafka is the core of the streaming service to ensure reliability, high availability, and scalability. All communications are transmitted using the Secure Sockets Layer (SSL) standard protocol, and data are encrypted at rest to ensure security. App access is controlled using a secure username-password combination.

**Discussion**

**Principal Findings**

Self-management of GDM is vital for controlling blood glucose levels and minimizing complications for both mother and baby [10,52]. In this paper, we described the design and development of GDM-DH, a culturally tailored GDM management app targeted for use among patients and health care providers in Nepal. Following the SCT framework [53], the GDM-DH app assists in self-management of GDM by enhancing the patients’ knowledge of and self-efficacy in adhering to blood glucose monitoring and recommended diet and physical activity regimens. With respect to the health care providers, the app’s web-based portal offers easy data visualization to track patient progress and treatment response, facilitating informed clinical decision-making at the point of care.

The growing literature highlights the importance of culturally tailoring health interventions, that is, adapting the intervention content and instructions according to the target users’ culture, diet, language, religion, customs, and beliefs [54-56]. Several studies [56-59], including systematic reviews [60], have found that culturally tailored programs and interventions are effective at improving disease knowledge, behavioral outcomes (e.g., physical activity), access to care, and clinical outcomes, including glycated hemoglobin (HbA1C) levels in patients with diabetes. Our GDM-DH app development incorporated a user-centered design approach that actively involved end users and used ethnographic and human-computer interaction methodologies to better understand and meet their needs. The user-centered design approach is especially paramount to developing culturally tailored mobile interventions, ensuring app engagement, and promoting digital health equity in low-income countries, such as Nepal [54].

Although mobile technology has been widely applied and proven efficacious for self-management of diabetes outside of pregnancy [29,32,41,61], mobile app–based lifestyle interventions for GDM are just emerging, even in high-income countries [62-68]. To date, there are only a few published randomized controlled trials that have evaluated mobile app–based solutions for GDM management [63-68]. Two recent reviews [69,70] on app-based interventions for GDM concluded that most existing studies were of moderate quality and were underpowered to detect effects on perinatal outcomes but, overall, indicated improved glycemic control in the mobile intervention groups compared to standard care alone. However, most existing app-based interventions for GDM management focus on remote blood glucose monitoring, with manual feedback from health care providers [24,69,70], which can be resource intensive and burdensome for both providers and participants, thus limiting the potential for widespread dissemination and impact, particularly in low-resource settings, such as Nepal. Additionally, despite evidence showing that lifestyle and T2D interventions based on behavior change theory are more effective [71-73], we found only 2 studies [67,68] that incorporated relevant theories in their mobile intervention for GDM. Furthermore, only 2 studies [65,67] involved input from target users during app design and development, which is critical for ensuring the effectiveness and acceptability of evidence-based interventions [74].

**Strengths**

Our GDM-DH app overcomes existing limitations and represents an advance over previous mobile interventions for GDM as it provides a comprehensive solution for GDM management without the need for additional work from health care professionals and incorporates user-centered principles and theory-based BCTs to meet the specific needs and technological literacy of our target users. To the best of our knowledge, the GDM-DH app is also the first to contain a social support component by including a social network feature. Although the educational content and custom food tracker (and visual aids for calorie/carbohydrate estimation) in our GDM-DH app were specifically designed for our target population of Nepalese women, they can be easily adapted and scaled to other contexts and populations by applying similar user-designed principles.

**Limitations**

Several limitations of the GDM-DH app and its development process are worth mentioning. First, the number of participants in our usability study was limited. Additionally, during usability testing, we may have observed the best-case scenario for comfort and confidence in using the app, leading us to overestimate the true usability and technological proficiency. Second, we did not use a structured framework, such as the Delphi method [75], to organize and structure the discussions to guide our GDM-DH app and intervention development, which will make it difficult for others to replicate our study procedures. Nonetheless, we used many of the elements of the Delphi framework, including an iterative approach, the use of experts, and group-based responses. Similarly, although we did not use a structured framework to guide the modification and optimization of the GDM-DH app and intervention content, our prespecified criteria, including feasibility, scalability, and affordability, align with existing frameworks designed to optimize and evaluate an intervention prior to implementation (e.g., multiphase optimization strategy, or MOST [76], and Acceptability, Practicability, Effectiveness, Affordability, Spill-Over Effects, and Equity, or APEASE [77], criteria). Due to resource limitations, we were unable to address all the needs and
suggestions provided by our target users. For instance, the current version of the app has limited social support, but future versions could incorporate features such as a discussion forum to foster a stronger network and support system among users. The manual entry of blood glucose and blood pressure levels is also a limitation. Since the app was designed to address the cultural barriers and technological literacy of a specific population, generalizability is a potential issue, and adapting the app to other settings and contexts would require a similar level of user research and testing among the target populations. Large-scale cluster-randomized clinical trials at multiple urban and rural sites in Nepal are needed to establish the effectiveness and generalizability of the GDM-DH app to women with GDM across the country.

Future Research
In the future, we plan to further optimize the GDM-DH app by including Bluetooth-enabled data entry and advanced smartphone functionalities, such as multimedia push notifications, and gamification features, which have been shown to increase retention and improve engagement with mHealth interventions [78-80]. Push notifications enable on-the-go delivery of intervention content, providing the necessary trigger and reinforcement when the specific intervention is most needed or is most convenient for the user. Gamification (application of game elements) features provide entertainment and intrinsic/extrinsic motivation (eg, point scores, badges, levels, a leaderboard) to promote sustained engagement with the app [81-85].

Conclusion
The GDM-DH app targets specific needs identified by our target population in our pilot research and has unique features, including a social support feature, visual aids for carbohydrate estimation, and a comprehensive support system without imposing an additional provider burden. A proof-of-concept pilot clinical trial (NCT04198857) to study the feasibility, acceptability, and preliminary efficacy of the GDM-DH app is currently underway.

Acknowledgments
This project was supported by the National Institutes of Health/Fogarty International Center (grant #NIH 1R21TW011377-01).

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors’ Contributions
Study concept and design were managed by SR; acquisition, analysis, and interpretation of data, AS (Aarthi Shanmugavel), PS, AS (Archana Shrestha), JN, AS (Abha Shrestha), J-FD, and SR; drafting of the manuscript and literature review, AS (Aarthi Shanmugavel) and SR; critical revision of the manuscript for important intellectual content, AS (Aarthi Shanmugavel), PS, AS (Archana Shrestha), JN, AS (Abha Shrestha), J-FD, and SR; and statistical analysis, JN and SR. All authors have read and approved the final manuscript.

Conflicts of Interest
None declared.

References


85. MacPherson MM, Merry KJ, Locke SR, Jung ME. Effects of mobile health prompts on self-monitoring and exercise behaviors following a diabetes prevention program: secondary analysis from a randomized controlled trial. JMIR Mhealth Uhealth 2019 Sep 05;7(9):e12956 [FREE Full text] [doi: 10.2196/12956] [Medline: 31489842]

Abbreviations

BCT: behavior change technique
GDM: gestational diabetes mellitus
GDM-DH: Garbhakalin Diabetes athawa Madhumeha—Dhulikhel Hospital
GWG: gestational weight gain
HIPAA: Health Insurance Portability and Accountability Act
KII: key informant interview
mHealth: mobile health
MVP: minimum viable product
PEN: Package of Essential Noncommunicable
SCT: social cognitive theory
T2D: type 2 diabetes

Edited by A Mavragani; submitted 14.09.23; peer-reviewed by S Locke, L Garnweidner-Holme; comments to author 19.10.23; revised version received 11.12.23; accepted 14.12.23; published 17.01.24.

Please cite as:
Shanmugavel A, Shakya PR, Shrestha A, Nepal J, Shrestha A, Daneault JF, Rawal S
Designing and Developing a Mobile App for Management and Treatment of Gestational Diabetes in Nepal: User-Centered Design Study
JMIR Form Res 2024;8:e50823
URL: https://formative.jmir.org/2024/1/e50823
doi:10.2196/50823
PMID:38231562
Preoperative Virtual Reality to Expose Patients With Breast Cancer to the Operating Room Environment: Feasibility and Pilot Case Series Study

Jordana L Sommer1,2, MA; Kristin Reynolds3, PhD; Pamela Hebbard4,5, MD; Michael S D Smith6, MSc; Natalie Mota7, PhD; W Alan C Mutch1, MD; Jessica Maples-Keller8, PhD; Leslie Roos2, PhD; Renée El-Gabalawy1,2,3,5,7, PhD

1Department of Anesthesiology, Perioperative and Pain Medicine, University of Manitoba, Winnipeg, MB, Canada
2Department of Psychology, University of Manitoba, Winnipeg, MB, Canada
3Department of Psychiatry, University of Manitoba, Winnipeg, MB, Canada
4Department of Surgery, University of Manitoba, Winnipeg, MB, Canada
5CancerCare Manitoba, Winnipeg, MB, Canada
6National Research Council Canada, Winnipeg, MB, Canada
7Department of Clinical Health Psychology, University of Manitoba, Winnipeg, MB, Canada
8Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA, United States

Corresponding Author:
Jordana L Sommer, MA
Department of Anesthesiology, Perioperative and Pain Medicine
University of Manitoba
AE207 Harry Medovy House
671 William Avenue
Winnipeg, MB, R3E 0Z2
Canada
Phone: 1 204 787 4713
Email: sommerj@myumanitoba.ca

Abstract

Background: Clinically elevated preoperative distress and anxiety are common among patients undergoing cancer surgery. Preoperative interventions have been developed to mitigate this distress and anxiety but are inconsistent in efficacy and feasibility for broad implementation.

Objective: This preliminary pilot study aims to assess the feasibility and utility of a newly developed virtual reality (VR) intervention to expose patients awaiting breast cancer surgery to the operating room environment and a simulation of anesthetic induction.

Methods: Patients undergoing breast cancer surgery (N=7) were assigned to the VR intervention or control (treatment as usual) group and completed self-report measures of distress and anxiety before surgery, on the day of surgery, and after surgery (5 and 30 d postoperatively). Those in the intervention group trialed the VR simulation 1 to 2 weeks preoperatively and provided qualitative and quantitative feedback. We assessed the feasibility of recruitment capability and study design and evaluated participants’ impressions of the intervention using self-report rating scales and open-ended questions. We also descriptively examined distress and anxiety levels throughout the duration of the study.

Results: Recruitment occurred between December 2021 and December 2022 and progressed slowly (rate: 1 participant/7 wk on average; some hesitancy because of stress and being overwhelmed). All participants who consented to participate completed the entire study. All participants were female and aged 56 (SD 10.56) years on average. In total, 57% (4/7) of the participants were assigned to the intervention group. On average, intervention participants spent 12 minutes engaged in the VR simulation. In general, the intervention was rated favorably (eg, clear information, enjoyable, and attractive presentation; mean% agreement 95.00-96.25, SD 4.79-10.00) and as helpful (mean% agreement 87.50, SD 25.00). Participants described the intervention as realistic (eg, “It was realistic to my past surgical experiences”), impacting their degree of preparedness and expectations for surgery (eg,
“The sounds and sights and procedures give you a test run; they prepare you for the actual day”), and having a calming or relaxing effect (eg. “You feel more relaxed for the surgery”).

Conclusions: This preoperative VR intervention demonstrated preliminary feasibility among a sample of patients undergoing breast cancer surgery. Results and participant feedback will inform modifications to the VR intervention and the study design of a large-scale randomized controlled trial to examine the efficacy of this intervention.

Trial Registration: ClinicalTrials.gov NCT04544618; https://clinicaltrials.gov/study/NCT04544618

(JMIR Form Res 2024;8:e46367) doi:10.2196/46367

KEYWORDS
virtual reality; preoperative anxiety and distress; breast cancer surgery; anesthesia; feasibility; pilot

Introduction

Preoperative Distress

Preoperative distress, a situational emotional reaction (eg, fear, worry, and helplessness), is common among surgical patients [1,2]. Extant research has identified several adverse health-related perioperative outcomes of preoperative distress in both the presence and absence of a mental disorder [3-7]. In particular, patients undergoing cancer surgery experience clinically meaningful elevated rates of preoperative distress, ranging from 23% to 77% in recent research [8-10]. Preoperative distress is also associated with various adverse health-related outcomes for patients undergoing cancer surgery specifically, including increased postoperative pain, nausea, discomfort, and fatigue, among others [11-15]. In recognition of its detrimental impact, the National Comprehensive Cancer Network (NCCN) designated distress as the “6th vital sign” [16].

Preoperative Psychological Interventions

Receiving a cancer diagnosis is a significant and life-altering event, often intensified by the necessity for major surgical intervention and an uncertain health trajectory. In considering the adverse health implications of psychological distress, several preoperative interventions (eg, education, relaxation training, and stress management) have been developed that seek to improve psychological and physical functioning before surgery by establishing realistic expectations of the surgical process and helping patients cope with surgery-related uncertainty and distress [17-24]. However, the literature reveals conflicting evidence regarding the efficacy of many such interventions [17-23]. Importantly, the interventions that are supported by evidence require delivery by licensed health care providers [21,22] and often require multiple sessions, rendering them impractical for large-scale implementation, particularly within the constraints of a publicly funded health care system.

Virtual Reality Interventions

Virtual reality (VR) interventions have shown considerable promise in reducing psychological distress in nonsurgical contexts [25-30]. Research in this area has examined the effectiveness of VR exposure therapy for the treatment of anxiety and posttraumatic stress disorder [25,27,28,31-33], where the user virtually and systematically confronts feared content to overcome anxiety. Patients often prefer using VR for exposure therapy over traditional in vivo exposure [34,35], and it may also be more straightforward to administer. This innovative technology has also been gaining popularity in broader medical contexts and has shown promising results in pain management [36-40] and cognitive and physical rehabilitation among various medical populations [37,41-43]. In contrast to therapist-guided VR exposure used in mental health settings, which may be used as a component of one-on-one psychotherapy over a duration of months, VR interventions developed for use in medical settings do not typically require a specialized health care professional to administer and can often achieve desirable outcomes following a shorter duration of use [37,44-47].

Preparatory Interventions for Stress Exposure

In a preoperative context, VR could be used to psychologically intervene before patients develop clinically elevated distress and are affected by the adverse downstream effects of distress (while also targeting any existing distress about surgery). This is similar to stress inoculation training, a form of cognitive behavioral intervention, aimed at psychologically preparing individuals for future exposure to a stressful environment through preliminary exposure to elements of that environment [48]. This form of intervention has been adapted using VR technology [48-52], and preliminary evidence supports reductions in predeployment distress for military personnel using such interventions to prepare for combat [50,53,54]. In fact, similar methods have been applied to psychologically prepare patients before surgery, including operating room (OR) tours before surgery [55], given that the OR environment is noted as distressing for many surgical patients [3,56,57]. Although this intervention was associated with reductions in preoperative distress [55], it has limited feasibility for broad administration because of the infrequent availability of ORs for such purposes and the limited resources and personnel to implement this intervention.

Preoperative Applications of VR

The use of VR to expose patients to the OR environment and preoperative process resolves some of these limitations. A few studies have implemented such interventions to target preoperative distress and other perioperative outcomes, largely among pediatric patients (all but one of the identified studies) undergoing variable types of surgeries (eg, general, neurological, and plastics or ear, nose, and throat) [58-64]. Small-scale meta-analyses examining this literature support the initial efficacy of such interventions in reducing preoperative distress [65-67], although some studies have used VR distraction interventions (eg, using games or relaxation) as opposed to
exposing users to the OR environment. Importantly, this research is in its infancy, and only a few studies exist in this area to date, supporting the need for further exploration.

Gaps in the Literature
Although the preoperative VR interventions described in the preceding section demonstrate preliminary efficacy in mitigating preoperative distress and potential for broad implementation within the constraints of our health care system (ie, relatively low cost, do not require specialized professional training to administer, can be used repeatedly in different settings, translated into multiple languages, and adapted across surgery types), studies examining these interventions are not without limitations. First, most studies in this area have focused on samples of pediatric patients undergoing surgery; further research is needed to establish the efficacy of such interventions among adult patient samples. Second, no identified studies to date have evaluated a preoperative VR intervention using patients scheduled to undergo an oncological procedure, a population with elevated levels of preoperative distress [8-10]. Third, existing preoperative VR interventions have limited immersion capabilities (eg, lack of user embodiment [ie, the ability to visualize and manipulate virtual representations of the user’s body] and use of prerecorded virtual videos as opposed to a fully immersive virtual environment), which may weaken their impact on mitigating distress through reduced realism. Fourth, these studies lacked follow-up data beyond the acute postoperative phase (eg, <1 wk after discharge), which is needed to understand the long-term impact on postoperative recovery. Finally, many of these studies did not gather user feedback on the intervention, which is vital to help maximize the potential impacts of these interventions.

This Study
In light of these identified gaps, this study aims to assess the feasibility of, and preliminarily pilot (in case series format), a novel VR OR simulation targeting preoperative distress and anxiety among a sample of patients undergoing breast cancer surgery. Specifically, regarding feasibility, the aims are to assess recruitment capability and identify resulting sample characteristics, understand participants’ impressions of the study design and intervention, and evaluate data collection procedures and outcome measures. Finally, this study will also pilot-test the preliminary impact of the intervention on perioperative distress and anxiety in a case series format. The results of this study will inform modifications to the VR simulation and the design of a large-scale randomized controlled trial (RCT) to evaluate the efficacy of this intervention.

Methods
Overview
This study used a single-blind randomized design to assess the feasibility of and pilot the VR simulation to expose patients undergoing breast cancer surgery to the OR and preoperative process. This study represents an in-depth preliminary analysis of a larger pilot study (ClinicalTrials.gov; NCT04544618). Participants were assigned to the VR intervention group or the treatment-as-usual (control) group at the time of recruitment. Randomization was stratified according to the type of breast cancer surgery (with vs without reconstruction) and whether neoadjuvant chemotheraphy was received to enable equal proportions of participants with these characteristics across study groups; research demonstrates differences in distress levels according to these factors [68]. All participants completed self-report measures 1 to 2 weeks before surgery (ie, baseline; VR group participants tested the intervention at this time), on the day of surgery, 5 days after surgery, and 30 days after surgery. Notably, the initial planned design included a third study arm (ie, non–surgery-related VR; Nature Treks), which was ultimately dropped because of recruitment challenges. Ethical amendments were approved supporting this change (and others noted in the Recruitment Capability and Sample Characteristics section), and the trial registry has been updated accordingly.

Participants
We recruited a convenience sample of adult patients undergoing breast cancer surgery by describing the study at patients’ surgical oncology appointments and preoperative information sessions and circulating study posters. Interested patients provided their contact information to enable a telephone discussion with the research coordinator and eligibility screening (those viewing the poster contacted the coordinator directly). Participants were eligible if they were being scheduled to undergo breast cancer surgery under general anesthesia at the Health Sciences Centre (a tertiary care hospital in Winnipeg, Canada) and could speak and read English. Those unable to provide informed consent or unable to participate in a VR intervention (eg, owing to visual or auditory impairment) were excluded. Our initial target was to recruit 15 participants per group, with a study aim to evaluate recruitment capability.

Procedure
Participants randomized to the VR group trialed the intervention 1 to 2 weeks before surgery (baseline). Those in the control group received no additional intervention beyond their standard medical appointments and optional preoperative information sessions (offered to all patients). All participants completed self-report measures at baseline (those in the VR group received additional measures to assess intervention feedback). On the day of surgery, preoperative distress and anxiety were reassessed while the participants were in the preoperative holding area and again in the OR before anesthetic induction. At 5 days and 30 days after the operation, all participants were readministered a subset of the baseline measures, and those in the VR group provided additional intervention feedback at the 5-day postoperative assessment. The participants in the VR group completed baseline measures in person (at the time of the intervention), and all participants completed the day-of-surgery measures in person. All other measures were completed online through the web-based survey platform, Qualtrics (Qualtrics International Inc).

Intervention
A VR development team at the National Research Council of Canada, in collaboration with coauthors (RE and JLS), developed the VR OR simulation for use in this research (a
technical paper describing the simulation more in depth is in progress). The simulation development stages included creating an initial prototype based on the observation of surgeries and consultation with medical personnel, developing an anesthetic induction script based on example scripts provided by several anesthesiologists, integrating input from an anesthesiologist (WACM) on the initial prototype, and refining the platform through multiple iterations of feedback from coauthors. The VR simulation begins with the participant sitting on an examination table (reflected as the OR table in the simulation), wearing the VR head-mounted display, and holding the controllers (enabling user embodiment and visualization and manipulation of virtual arms). The participant is instructed to imagine it is their day of surgery, including how they might be feeling that day. The participant then spends at least 5 minutes exploring the virtual OR, which includes relevant machinery and equipment, personnel, sounds, and details such as a mammogram displayed on a computer screen. This free exploration is followed by a scripted portion, where the participant is instructed to lie supine on the virtual OR table and is taken through a mock anesthetic induction process led by the virtual anesthesiologist and nurse; the patient is prompted to answer questions similar to those they will be asked on the day of surgery (eg, name, date of birth, type of surgery, and allergies) and is virtually prepared with monitoring devices by the nurse (eg, blood pressure cuff, a pulse oximeter, and electrocardiogram stickers and electrodes). The simulation ends after the virtual oxygen mask is placed on the patient’s mouth and the screen darkens (refer to Multimedia Appendix 1). We used the Oculus Rift S VR system (Meta Platforms) with a cable connection to a laptop computer for the intervention administration.

Ethical Considerations

This study was approved by the University of Manitoba Health Research Ethics Board (#HS23957). All participants provided written informed consent before participation. No participant-identifying information was included with the study data. Each participant was assigned a study identification number, which was used to collate participant data over the study duration. All participants were provided with a CAD $25 (US $18.94) gift card after completing the study, and the cost of parking for those attending an intervention appointment was reimbursed.

Measures

Preoperative Distress

A total of 4 self-report measures assessed preoperative distress, including 2 preoperative-specific scales (Preoperative Intrusive Thoughts Inventory [PITI] and Amsterdam Preoperative Anxiety Information Scale [APAIAS]) and 2 nonspecific visual analog scales (NCCN Distress Thermometer and adapted “Anxiety Thermometer”). The PITI and APAIAS were only administered at baseline and on the day of surgery, whereas the Distress and Anxiety Thermometers were assessed at all 4 time points (and in the OR before induction). At the 5-day postoperative follow-up, the participants were asked which measure best captured their experience of distress or anxiety, which will inform the selection of the primary outcome measure for the upcoming RCT. At 5-day postoperative follow-up, they were also prompted to retrospectively rate their level of distress/anxiety from 0 (no distress/anxiety) to 10 (extreme distress/anxiety) corresponding to 8 different “events” ranging from prediagnosis (average level of distress/anxiety before receiving a cancer diagnosis) until the 5-day follow-up.

The PITI is a validated and reliable 20-item self-report measure of preoperative anxiety [69]. Items (eg, “I worry that I won’t wake up”) are rated on a 4-point scale, ranging from 0 (not at all) to 3 (most of the time). Summing the items yields a total score ranging from 0 to 60, where a score ≥15 indicates clinically significant preoperative anxiety [69]. The APAIAS is a validated and reliable 6-item measure of preoperative anxiety [70]. Items (eg, “I am worried about the procedure”) are rated on a 5-point scale, ranging from 1 (not at all) to 5 (extremely). A total score is calculated by summing all items, ranging from 6 to 30, where a score ≥10 indicates clinically elevated preoperative anxiety [70]. The NCCN Distress Thermometer is a visual analog scale that resembles a thermometer, with a scale ranging from 0 (no distress) to 10 (extreme distress) [71]. This has been validated among several oncology samples [72,73]. Distress is rated using a “past-week” timeframe (modified to present time in the OR), and a cut-off score of 4 indicates clinically elevated distress [73]. Because of the common interchangeable use of the terms distress and anxiety within the perioperative and oncology literature and the lack of clear differentiation between these terms, we adapted the Distress Thermometer to create an Anxiety Thermometer.

VR Impressions and Feedback

Participants provided self-reported feedback on the VR simulation at 2 different time points. Feedback measures were developed in accordance with previous research [74]. After trialing the intervention, the participants were provided with a list of statements about their experience using the simulation (eg, “I found the VR intervention was helpful”), which they rated from 0% (completely disagree) to 100% (completely agree). The participants are also asked whether they experienced any motion sickness during the intervention (0 [none] to 3 [severe]), followed by open-ended questions prompting intervention impressions (eg, what they liked or disliked) and whether they found the intervention worthwhile. Finally, the participants were asked about additional elements they wished to be included in the intervention and were offered multiple response options for selection (eg, being wheeled into the OR). At the 5-day postoperative follow-up, the VR participants provided additional intervention feedback (eg, overall impressions). The participants are then asked if or how they think the intervention impacted their surgery or recovery, whether they disliked anything about it, and if they have any other suggestions for improvements. The participants are prompted to elaborate on their responses to these questions.

Presence

The iGroup Presence Questionnaire [75] assessed the presence associated with the VR intervention at baseline, defined as the sense of being in the virtual environment. This valid and reliable (Cronbach α=.87) self-report measure is comprised of 14 items (eg, “I had a sense of acting in the virtual space, rather than
operating something from outside”), which are rated on a 7-point scale (−3 [fully disagree] to +3 [fully agree]). Items are summed to create 3 subscale scores (spatial presence, involvement, and realism), where higher scores indicate increased presence in the virtual environment.

**Sample Characteristics**

Participants self-reported their sociodemographic characteristics and health history at baseline, including age (assessed continuously), sex (female or male), marital status (single, married or common law, divorced or separated, or widowed), highest level of education (high school or less or some college or higher), stage of breast cancer, type of breast cancer surgery (eg, lumpectomy or single or double mastectomy with or without reconstruction), whether they are receiving chemotherapy before surgery, history of prior surgeries, mental health service use since receiving their cancer diagnosis, and history of receiving a mental health diagnosis. Various other self-report measures were administered throughout the study (eg, assessing depression, coping, and quality of life) to determine their feasibility for inclusion in the upcoming RCT (by calculating the proportion of missing data).

**Analytic Strategy**

Descriptive statistics assessed the consent rate, recruitment speed, attrition rate, and sample characteristics. We calculated the participation rate among the intervention and control groups, and we assessed quantitative and qualitative intervention feedback descriptively. We then calculated the proportion of missing data, and descriptive statistics determined which measure of preoperative distress or anxiety was rated most favorably. Finally, we presented participants’ levels of distress and anxiety across the perioperative period (baseline to 30-day follow-up) descriptively in a case series format.

**Results**

**Feasibility Aims**

**Recruitment Capability and Sample Characteristics**

Recruitment was initiated on December 1, 2021. Between initiation and December 1, 2022, a total of 14 prospective participants were identified (n=5, 36% were identified in the final 2 months of recruitment). Of these 14 individuals, 12 (86%) contacted the study coordinator directly, 1 (7%) had their information provided by a health professional, and 1 (7%) expressed interest while attending a preoperative information session (in-person sessions were suspended until November 2022). Of these 14 individuals, 7 (50%) consented and participated, 4 (29%) were ineligible (eg, required to isolate until surgery or already had surgery), and 3 (21%) withdrew after providing verbal consent but before providing written consent (reasons: n=1, 33% too many appointments and unable to focus on anything else; n=1, 33% unwilling to come in person to try the VR; and n=1, 33% overwhelmed with family responsibilities; n=2, 67% had been randomized to the initial third arm before dropping out). The approximate recruitment speed for those who consented was 1 participant every 7 weeks, on average. In total, 57% (4/7) of the participants were assigned to the intervention group. Of those who provided written informed consent, 100% (7/7) completed the study. Because of ongoing recruitment challenges, the study target population was broadened 5 months after the initiation of recruitment to include any patients undergoing cancer surgery, as opposed to patients undergoing breast cancer surgery only. To date, no patients undergoing non-breast cancer surgery have expressed an interest in participating.

The participants were aged 56.43 (SD 10.56) years on average, and all were female. The participants were most commonly married (3/7, 43%), and the majority (5/7, 71%) had some college education or higher. The breast cancer stage of patients was most commonly uncertain or unknown (4/7, 57%). The most common surgical procedure was lumpectomy (4/7, 57%), and 43% (3/7) of the participants were planning to undergo reconstructive surgery. Most participants (6/7, 86%) had not received chemotherapy before their surgery, and most participants (6/7, 86%) had ≥1 prior surgeries. A total of 57% (4/7) of the participants reported receiving a mental health diagnosis in their lifetime (depression, anxiety, or substance use disorder), and 29% (2/7) of the participants indicated that they sought professional mental health support after receiving their cancer diagnosis. Most participants (4/7, 57% to 6/7, 86%) had clinically elevated preoperative distress or anxiety at baseline (Table 1).

See Table 1 for descriptive statistics of sample characteristics.
Table 1. Sample characteristics for patients undergoing breast cancer surgery participating in the feasibility and pilot study to evaluate preoperative virtual reality (n=7).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y), mean (SD)</td>
<td>56.43 (10.56)</td>
</tr>
<tr>
<td>Sex (female), n (%)</td>
<td>7 (100)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>2 (29)</td>
</tr>
<tr>
<td>Married or common law</td>
<td>3 (43)</td>
</tr>
<tr>
<td>Divorced or separated</td>
<td>2 (29)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>High school or less</td>
<td>2 (29)</td>
</tr>
<tr>
<td>Some college or higher</td>
<td>5 (71)</td>
</tr>
<tr>
<td><strong>Stage of breast cancer, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Uncertain (0-1, 1-2, or other unknown)</td>
<td>4 (57)</td>
</tr>
<tr>
<td>Stage 1</td>
<td>1 (14)</td>
</tr>
<tr>
<td>Stage 2</td>
<td>2 (29)</td>
</tr>
<tr>
<td><strong>Type of surgery, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Lumpectomy</td>
<td>4 (57)</td>
</tr>
<tr>
<td>Single mastectomy without reconstruction</td>
<td>1 (14)</td>
</tr>
<tr>
<td>Single mastectomy with immediate reconstruction</td>
<td>1 (14)</td>
</tr>
<tr>
<td>Double mastectomy with immediate reconstruction</td>
<td>1 (14)</td>
</tr>
<tr>
<td>Undergoing reconstruction, n (%)</td>
<td>3 (43)</td>
</tr>
<tr>
<td>Neoadjuvant chemotherapy, n (%)</td>
<td>1 (14)</td>
</tr>
<tr>
<td>History of prior surgery, n (%)</td>
<td>6 (86)</td>
</tr>
<tr>
<td>Sought professional mental health support since cancer diagnosis, n (%)</td>
<td>2 (29)</td>
</tr>
<tr>
<td>Lifetime mental health diagnosis, n (%)</td>
<td>4 (57)</td>
</tr>
<tr>
<td><strong>Clinically significant preoperative distress or anxiety at baseline, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>PITI⁸</td>
<td>4 (57)</td>
</tr>
<tr>
<td>APAIS⁹</td>
<td>6 (86)</td>
</tr>
<tr>
<td>Distress thermometer</td>
<td>4 (57)</td>
</tr>
<tr>
<td>Anxiety thermometer</td>
<td>5 (71)</td>
</tr>
<tr>
<td>Intervention group, n (%)</td>
<td>4 (57)</td>
</tr>
</tbody>
</table>

⁸PITI: Preoperative Intrusive Thoughts Inventory.
⁹APAIS: Amsterdam Preoperative Anxiety Information Scale.

**Participant Impressions of the Study Design and Intervention**

All participants assigned to the control group completed the entire study. All participants assigned to the intervention group, who provided written informed consent, tested the intervention within 2 weeks before their surgery and completed the study. The participants in the intervention group spent 12 minutes engaged in the simulation, on average, and reported variable levels of presence while trialing the VR simulation (spatial presence: mean 8.75, involvement: mean 0.75, and realism: mean −2.50; refer to Table 2 for maximum ranges); the participants reported having a sense of being physically present in the virtual environment, with only partial attention devoted to the virtual environment, and moderate ratings of realism.
Table 2. Quantitative intervention impressions at baseline for patients undergoing breast cancer surgery assigned to the intervention group for the feasibility and pilot study evaluating preoperative virtual reality (VR; n=4).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>VR duration (min), mean (SD)</td>
<td>11.64 (1.08)</td>
</tr>
<tr>
<td>Presence: spatial presence subscale (maximum range: −15 to 15), mean (SD)</td>
<td>8.75 (3.30)</td>
</tr>
<tr>
<td>Presence: involvement subscale (maximum range: −12 to 12), mean (SD)</td>
<td>0.75 (1.26)</td>
</tr>
<tr>
<td>Presence: realism subscale (maximum range: −15 to 15), mean (SD)</td>
<td>−2.50 (3.70)</td>
</tr>
<tr>
<td>The way information was presented was clear and understandable (0%–100%), mean (SD)</td>
<td>95.00 (10.00)</td>
</tr>
<tr>
<td>I enjoyed my session with the VR program (0%–100%), mean (SD)</td>
<td>96.25 (4.79)</td>
</tr>
<tr>
<td>I could understand and act on the information provided by the program (0%–100%), mean (SD)</td>
<td>93.75 (7.50)</td>
</tr>
<tr>
<td>The program had a very attractive presentation (0%–100%), mean (SD)</td>
<td>95.00 (5.77)</td>
</tr>
<tr>
<td>I had to look for assistance when I used this program (0%–100%), mean (SD)</td>
<td>42.50 (43.49)</td>
</tr>
<tr>
<td>The VR program froze or stopped unexpectedly (0%–100%), mean (SD)</td>
<td>5.00 (10.00)</td>
</tr>
<tr>
<td>I found the VR intervention was helpful (0%–100%), mean (SD)</td>
<td>87.50 (25.00)</td>
</tr>
<tr>
<td>The VR intervention eased my anxiety/concerns about the OR(^a) (0%–100%), mean (SD)</td>
<td>55.00 (47.96)</td>
</tr>
<tr>
<td>The VR intervention eased my anxiety/concerns about the anesthesia (0%–100%), mean (SD)</td>
<td>60.00 (45.46)</td>
</tr>
<tr>
<td>The VR intervention eased my anxiety/concerns about my surgery (0%–100%), mean (SD)</td>
<td>46.25 (38.16)</td>
</tr>
<tr>
<td>The VR intervention worsened my anxiety/concerns about the OR (0%–100%), mean (SD)</td>
<td>37.50 (47.87)</td>
</tr>
<tr>
<td>The VR intervention worsened my anxiety/concerns about the anesthesia (0%–100%), mean (SD)</td>
<td>25.00 (28.87)</td>
</tr>
<tr>
<td>The VR intervention worsened my anxiety/concerns about my surgery (0%–100%), mean (SD)</td>
<td>30.00 (51.96)</td>
</tr>
<tr>
<td>Experienced motion sickness, n (%)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Participating in the VR intervention was worthwhile considering time commitment, n (%)</td>
<td>4 (100)</td>
</tr>
<tr>
<td><strong>Other elements you would have liked to included, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Being wheeled into the OR</td>
<td>1 (25)</td>
</tr>
<tr>
<td>Try on equipment (eg, oxygen mask) while engaged in the simulation</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Learn about the various machines I saw in the OR</td>
<td>1 (25)</td>
</tr>
<tr>
<td>Ask the virtual anesthetist or nurse questions about my surgery</td>
<td>1 (25)</td>
</tr>
<tr>
<td>None of the above</td>
<td>2 (50)</td>
</tr>
<tr>
<td>Other (“real time pulse/heart rate”)</td>
<td>1 (25)</td>
</tr>
</tbody>
</table>

\(^a\)OR: operating room.

There were minor technical difficulties for all 4 participants during the simulation (eg, difficulty finding the correct position lying down when prompted by the VR nurse), and the program needed to be restarted midway for 2 of the participants. In general, the participants found that the intervention presented information clearly, was enjoyable, easy to understand, and had an attractive presentation (mean\(\%\) agreement range: 93.75–96.25, SD range: 4.79–10.00). It was also generally rated as helpful (mean\(\%\) agreement 87.50, SD 25.00), and all participants considered participating in the VR intervention worthwhile considering the time commitment. The participants gave mixed ratings at baseline regarding the impact of the intervention on anxiety and concerns about the OR, anesthesia, and surgery. Given a list of suggestions for elements to be added to the intervention, a single participant selected each of the following: (1) being wheeled into the OR, (2) learn about the machines I saw in the OR, (3) ask the virtual anesthesiologist or nurse questions about surgery, and (4) other: “real time pulse/heart rate.”

Regarding open-ended feedback, multiple participants commented on the realism of the intervention, the impact of the intervention on expectation formation regarding surgery, and the calming or relaxing effect of the intervention (Textbox 1).
Textbox 1. Qualitative intervention impressions for patients undergoing breast cancer surgery assigned to the intervention group for the feasibility and pilot study evaluating preoperative virtual reality (VR; n=4).

**Open-ended feedback (at baseline)**

- What did you like about the VR intervention
  - “It is very realistic”
  - “It was realistic to my past surgical experiences, it was interactive and could play a bit with it”
  - “What to expect”
  - “Just getting the feel of an OR” [operating room]

- What did you dislike about the VR intervention
  - “Nothing”
  - “Scary”
  - “The program calibrated my body position a few times and had to be reset which is why I was more present in the real world than in the VR world”
  - “Seemed like I was waiting for an hour until it told me to lie down”

- If you found it helpful: in what ways was the VR intervention helpful
  - “The sounds and sights and procedures give you a test run- prepares you for the actual day”
  - “Yes”
  - “Was helpful in that it reminded me of all the noises, lights, and people necessary in an OR”
  - “You feel more relaxed for the surgery”

- Explain why it was or was not worthwhile
  - “Gave me information and made me think of my feelings, made me feel better”
  - “I like to help with research and I’m curious about VR and mental health initiatives”
  - “Knowing what to expect”
  - “Think I can relax a bit now when it’s time for me to have my surgery”

**Assessed 5 d after surgery**

- Overall impression of the VR intervention
  - “It was very good, very real to life. I liked it”
  - “Head set didn’t work well”
  - “I had past surgery and it was familiar from memory and with current surgery experience”
  - “I thought it was a good way to help calm some of my fears”

- Elements from the OR that were missing from the VR which would have been helpful to include
  - “No I think they covered everything”
  - “If when they are putting stickers on etc. you would maybe lightly touch the spot”
  - “Not that I remember. I wasn’t paying much attention to what everyone was doing or the equipment”
  - “More condensed room, just focus on the 2 people in your face”

- Images or experiences from the VR intervention that stuck with you following the intervention
  - “No”
  - “The mask at the end”
  - “The nurse moved in on my too quickly and startled me because she was so close so suddenly”
  - “The lights”

- Components of in-hospital experiences on day of surgery that would have been helpful to include in VR
“The waiting in the surgical admitting area. Sitting for a long time in a chair in the gown with IV pick in”

“I didn’t get to wake up in the VR but it may be cool to wake up. You aren’t alone when they wake you up in case that unknown freaks people out”

“The actual experience happened a lot quicker than the virtual experience. Speed up the simulation”

How, if at all, do you think the VR simulation impacted your surgery or recovery

“It makes a person more relaxed in the operating room”

“Once they had me in the surgery room it was very fast”

“If I didn’t have past experience then it would have helped me a lot but I was already familiar”

“I believe it assisted me in that I was able to see the inside of an OR”

Was there anything you disliked about the VR intervention (if yes, please describe)

“No”

“Sadly the program had to be reset a bunch of times because...the orientation was off. It brought me out of it”

“It was way too long just sitting there and waiting for something to happen”

Suggestions regarding how we can improve the VR simulation

“No it was very informative”

“Have the room smaller and things not so far away. People need to be closer to you”

Data Collection Procedures and Outcomes Measures

Across all the time points, only 0.7% of the data were missing. Most participants (5/7, 71%) reported that, of the different measures assessing anxiety and distress, the PITI best captured their experiences.

Table 3 outlines the sample characteristics, and Table 4 outlines the perioperative levels of distress and anxiety of the participants in the control and intervention groups.

Table 3. Participant characteristics.

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>Group</th>
<th>Age (y)</th>
<th>Current surgery</th>
<th>Prior surgery</th>
<th>Mental health history</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Control</td>
<td>60s</td>
<td>Lumpectomy</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>P2</td>
<td>Control</td>
<td>60s</td>
<td>Lumpectomy</td>
<td>Single mastectomy (&gt;10 y ago)</td>
<td>Depression</td>
</tr>
<tr>
<td>P3</td>
<td>Control</td>
<td>40s</td>
<td>Double mastectomy with immediate reconstruction</td>
<td>Broken arm and appendectomy</td>
<td>None</td>
</tr>
<tr>
<td>P4</td>
<td>Intervention</td>
<td>60s</td>
<td>Lumpectomy</td>
<td>“Replacements” and “abnormal cell removals”</td>
<td>Mental health leave (no diagnosis)</td>
</tr>
<tr>
<td>P5</td>
<td>Intervention</td>
<td>40s</td>
<td>Single mastectomy with immediate reconstruction</td>
<td>Thyroid surgery &gt;5 y ago</td>
<td>None</td>
</tr>
<tr>
<td>P6</td>
<td>Intervention</td>
<td>50s</td>
<td>Lumpectomy</td>
<td>Lumpectomy, fibroids removed, hysterectomy, cervix and ovaries removed, and deviated septum repair</td>
<td>Depression and anxiety</td>
</tr>
<tr>
<td>P7</td>
<td>Intervention</td>
<td>50s</td>
<td>Single mastectomy without reconstruction</td>
<td>Arm and cesarean section</td>
<td>Depression and substance use disorder</td>
</tr>
</tbody>
</table>

Age range.
Table 4. Perioperative distress or anxiety for patients ongoing breast cancer surgery participating in the feasibility and pilot study to evaluate preoperative virtual reality.

<table>
<thead>
<tr>
<th>Participant ID, group, and measure</th>
<th>Perioperative distress or anxiety&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Clinically elevated&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
<td>Preoperative</td>
</tr>
<tr>
<td><strong>P1: control group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PITI&lt;sup&gt;d&lt;/sup&gt; total</td>
<td>13.00</td>
<td>7.00</td>
</tr>
<tr>
<td>APAIS&lt;sup&gt;d&lt;/sup&gt; total</td>
<td>10.00</td>
<td>12.00</td>
</tr>
<tr>
<td>Distress thermometer</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Anxiety thermometer</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>P2: control group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PITI total</td>
<td>26.00</td>
<td>32.00</td>
</tr>
<tr>
<td>APAIS total</td>
<td>13.00</td>
<td>18.00</td>
</tr>
<tr>
<td>Distress thermometer</td>
<td>3.00</td>
<td>6.50</td>
</tr>
<tr>
<td>Anxiety thermometer</td>
<td>3.00</td>
<td>6.50</td>
</tr>
<tr>
<td><strong>P3: control group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PITI total</td>
<td>42.00</td>
<td>37.00</td>
</tr>
<tr>
<td>APAIS total</td>
<td>19.00</td>
<td>17.00</td>
</tr>
<tr>
<td>Distress thermometer</td>
<td>8.00</td>
<td>9.00</td>
</tr>
<tr>
<td>Anxiety thermometer</td>
<td>9.00</td>
<td>9.00</td>
</tr>
<tr>
<td><strong>P4: intervention group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PITI total</td>
<td>10.00</td>
<td>12.00</td>
</tr>
<tr>
<td>APAIS total</td>
<td>11.00</td>
<td>8.00</td>
</tr>
<tr>
<td>Distress thermometer</td>
<td>6.00</td>
<td>3.00</td>
</tr>
<tr>
<td>Anxiety thermometer</td>
<td>6.00</td>
<td>9.00</td>
</tr>
<tr>
<td><strong>P5: intervention group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PITI total</td>
<td>47.00</td>
<td>40.00</td>
</tr>
<tr>
<td>APAIS total</td>
<td>21.00</td>
<td>24.00</td>
</tr>
<tr>
<td>Distress thermometer</td>
<td>10.00</td>
<td>10.00</td>
</tr>
<tr>
<td>Anxiety thermometer</td>
<td>10.00</td>
<td>9.00</td>
</tr>
<tr>
<td><strong>P6: intervention group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PITI total</td>
<td>11.00</td>
<td>10.00</td>
</tr>
<tr>
<td>APAIS total</td>
<td>9.00</td>
<td>10.00</td>
</tr>
<tr>
<td>Distress thermometer</td>
<td>3.00</td>
<td>2.00</td>
</tr>
<tr>
<td>Anxiety thermometer</td>
<td>4.00</td>
<td>3.00</td>
</tr>
<tr>
<td><strong>P7: intervention group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PITI total</td>
<td>43.00</td>
<td>49.00</td>
</tr>
<tr>
<td>APAIS total</td>
<td>21.00</td>
<td>22.00</td>
</tr>
<tr>
<td>Distress thermometer</td>
<td>7.00</td>
<td>8.00</td>
</tr>
<tr>
<td>Anxiety thermometer</td>
<td>7.00</td>
<td>9.00</td>
</tr>
</tbody>
</table>

<sup>a</sup>Values represent total scores on each measure at each time point.

<sup>b</sup>Values represent the number of times a score is clinically elevated across the total number of measurements.

<sup>c</sup>OR: operating room.

<sup>d</sup>PITI: Preoperative Intrusive Thoughts Inventory.

<sup>e</sup>N/A: not applicable; PITI and Amsterdam Preoperative Anxiety Information Scale are specific to the preoperative period and were not administered.
in the OR or during the postoperative phase.

\footnote{APAIS: Amsterdam Preoperative Anxiety Information Scale.}

\footnote{Missing data because of surgery scheduling change.}

**Retrospective Reports of Distress or Anxiety**

As shown in Figure 1, among the control group, ratings of distress/anxiety remained stable (P1) or increased (P2 and P3) between baseline (within 2-wk preoperatively) and being in the OR on the day of surgery. Among the intervention group, ratings decreased between baseline (within 2-wk preoperatively; when VR was administered) and being in the OR for 50% (2/4; P4 and P5) of the participants.

![Figure 1. Retrospective reports of distress/anxiety among the control group and intervention group for patients undergoing breast cancer surgery participating in the feasibility and pilot study to evaluate preoperative virtual reality (VR). Blue guidelines outline the period between when the intervention group trialed the VR and participants’ day of surgery.](image)

**Discussion**

**Overview**

To our knowledge, this is the first study to date to examine the feasibility of, and preliminarily pilot, a novel preoperative VR intervention exposing patients undergoing breast cancer surgery to the OR and preoperative process. Overall, despite some recruitment challenges, the intervention was generally rated favorably and described, on average, as 87.5% (SD 25.00%) helpful by participants. The results of this study will inform modifications made to the VR intervention and the study design of an upcoming RCT evaluating this intervention.

The newly developed VR intervention exposed patients undergoing breast cancer surgery to the OR environment (including machinery, sounds, personnel, and other medical items [eg, surgical tools and mammogram]) and preoperative process (from being seated on the OR table until completion of anesthetic induction). The simulation was developed to mimic the real-life OR and preoperative experience based on a large tertiary care hospital in Winnipeg, Manitoba. Compared with other recently developed preoperative VR interventions [58-64], this was designed to be more immersive through the integration of user embodiment (including visualization of one’s virtual body and real-time manipulation of virtual arms) and is one of the few interventions designed for adult use and the only such intervention developed and tested in Canada.

Although some technical difficulties arose during the intervention (eg, simulation needing to be restarted and slight delay in progression because of imprecise positioning of participant arm or body), likely detracting from immersion, the participants described the intervention as realistic and commented on its impact on feeling more prepared or knowing what to expect for surgery and feeling more relaxed or calm about their upcoming surgery. The participants also rated the intervention favorably in terms of enjoyment, clarity of information, attractiveness, and helpfulness. Although the sample size of this study limits our ability to establish trends regarding the impact of the intervention on distress and anxiety, the participants rated 46% to 60% agreement (SD range 38.16%-47.96%), on average, that the intervention eased their anxiety, and for half of the intervention group participants (2/4, 50%), retrospective ratings of distress/anxiety declined between trialing the intervention and being in the OR. Notably, the participants also rated 25% to 38% (SD range 28.87%-51.96%) agreement, on average, that the intervention worsened their anxiety (immediately postintervention), although they did not indicate this when providing feedback postoperatively. This may suggest the activation of the “fear structure” within the simulation, which is noted as an important component of anxiety-based exposure interventions [76].
Although preliminary data support the feasibility of the VR intervention, we encountered challenges regarding recruitment for the study. This may have been impacted by various factors including changes to surgical scheduling during the COVID-19 pandemic (noted in recent research on patients with cancer [77]), prospective participants’ reported feelings of being overwhelmed and stressed by their own health or other commitments, and a strained health care system resulting in reduced resources to support recruitment (including canceling in-person preoperative information sessions for 10 months during the recruitment period, where recruitment was planned to take place). As noted, recruitment began improving over the final 2 months of the recruitment period, wherein 80% (4/5) of the individuals who expressed interest in the study provided consent to participate. Although speculative, this may suggest an impact of the changing centrality of the pandemic on recruitment capability. Interestingly, most participants (6/7, 86%) had a history of prior surgeries, which could have resulted in an increased willingness to participate. It may be worthwhile to consider modifications to our recruitment poster (eg, including the rationale for the intervention) to entice participation from those who have not undergone prior surgery. The study design elements, including data collection, intervention engagement, and participant retention, appear feasible based on the current data.

**Strengths and Limitations**

Despite the strengths of this study, including the novel preoperative VR intervention integrating user embodiment, evaluation of the feasibility of this intervention in a population with elevated estimates of clinically significant distress [8-10], collection of qualitative and quantitative intervention feedback, and inclusion of 2 iterations of postoperative follow-up data (5 and 30 days postoperatively; to be evaluated in an upcoming larger study), this study is not without limitations. First, recruitment challenges limited our sample size for this initial study; however, these challenges provided important information regarding the feasibility of implementing a larger study in the future. Second, there were a few technical difficulties encountered when administering the VR intervention, detracting from user immersion. Finally, although not directly investigated, distress in this population (and assessed using nonspecific measures) is likely to be influenced by many factors in addition to surgery. This particular intervention may not be very beneficial or impactful for those with primarily non–surgery-related distress.

**Implications**

Importantly, these limitations, along with the data collected as part of this study, provide important insights to inform modifications to the intervention and study design before the implementation of a large-scale RCT to evaluate the efficacy of this intervention. Regarding recruitment, we will consider ways to target enhancing the involvement of health care professionals in spreading awareness of the study to potentially eligible patients while continuing to attempt recruitment at the newly reinstated in-person preoperative information sessions. In addition, we will consider including additional information about the intervention (and thus removing participant blinding) as part of the recruitment process. Changes to consider for the VR simulation include modifying requirements for the user’s body positioning to avoid unnecessary interruptions and potentially adding elements that participants noted would have been helpful (eg, the opportunity to learn about OR machines and ask questions to the virtual anesthetist or nurse). On the basis of participant feedback, this intervention has the potential to reduce levels of preoperative distress/anxiety by helping participants form more realistic expectations of the day of surgery before their operation (thus potentially reducing their perception of threat associated with the preoperative experience and enhancing their perceived ability to cope with this stressor). In line with recommendations based on other VR exposure-based interventions [78], having repeated exposure to the simulation may enhance the potential impact on mitigating distress/anxiety. Thus, it may be beneficial to assess the utility of providing participants with a 2D “screen-capture” video recording of their VR trial to watch on their own device multiple times in between trialing the intervention and their surgery. This may be an important avenue for future research evaluating this intervention.

Overall, this study established the initial feasibility of a novel preoperative VR intervention to expose patients undergoing breast cancer surgery to the OR and anesthetic induction process. These results will inform the study design of an upcoming large RCT to further examine this intervention. Participant feedback supports the utility and acceptability of this intervention and will inform future adaptations to the simulation. If demonstrated as efficacious in upcoming research, this intervention has the potential to be adapted across multiple surgery types and implemented on a broad scale to help mitigate preoperative distress.

**Acknowledgments**

The authors acknowledge Gabrielle Logan and Bronwen Grocott for their contributions to this research and Catherine Proulx and Vincent Gagnon Shaigetz for their contributions to the development of the virtual reality software. This study received funding from the Canadian Institutes of Health Research Canada Graduate Scholarship-Doctoral Award, University of Manitoba Graduate Fellowship, University of Manitoba Sheu L Lee Family Scholarship in Oncology Research, Women’s Health Research Foundation of Canada Graduate Scholarship, National Register for Health Service Psychologists Doctoral Award (Sommer), University of Manitoba Start-Up Funds, and Tri-Agency New Frontier’s in Research Grant (El-Gabalawy).

**Data Availability**

Data access is restricted to protect the confidentiality of participants and in light of ongoing research to expand the data set.
Conflicts of Interest
None declared.

Multimedia Appendix 1
Virtual reality simulation screen captures.

References


Abbreviations

APAIS: Amsterdam Preoperative Anxiety Information Scale
NCCN: National Comprehensive Cancer Network
OR: operating room
PITI: Preoperative Intrusive Thoughts Inventory
RCT: randomized controlled trial
VR: virtual reality

Edited by A Mavragani; submitted 08.02.23; peer-reviewed by W Trapp; comments to author 27.10.23; revised version received 16.11.23; accepted 04.12.23; published 17.01.24.

Please cite as:
Preoperative Virtual Reality to Expose Patients With Breast Cancer to the Operating Room Environment: Feasibility and Pilot Case Series Study
JMIR Form Res 2024;8:e46367
URL: https://formative.jmir.org/2024/1/e46367
doi:10.2196/46367
PMID:38231570
Educational Video Intervention to Improve Health Misinformation Identification on WhatsApp Among Saudi Arabian Population: Pre-Post Intervention Study

Ebtihal Alsaad1,2*, DDS, MSc; Sharifah AlDossary1,2*, MSc, PhD
1Department of Health Informatics, College of Public Health and Health Informatics, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia
2King Abdullah International Medical Research Center, Riyadh, Saudi Arabia
*all authors contributed equally

Corresponding Author:
Ebtihal Alsaad, DDS, MSc
Department of Health Informatics
College of Public Health and Health Informatics
King Saud bin Abdulaziz University for Health Sciences
Al Haras Al Watani St
Prince Mutib Ibn Abdullah Ibn Abdulaziz Rd, Ar Rimayah, Riyadh
Riyadh, PO Box 3660/ 11481
Saudi Arabia
Phone: 966 114299999
Email: ebtihalalsaad@gmail.com

Abstract

Background: Health misinformation can adversely affect individuals’ quality of life and increase the risk of mortality. People often fail to assess the content of messages before sharing them on the internet, increasing the spread of misinformation. The problem is exacerbated by the growing variety of digital information environments, especially social media, which presents as an effective platform for spreading misinformation due to its rapid information-sharing capabilities. Educational interventions have been developed to help consumers verify the validity of digital health information. However, tools designed to detect health misinformation on social media content have not been validated. Given the increased use of social media platforms, particularly WhatsApp, it is crucial to develop tools to help consumers assess the credibility of messages and detect misinformation.

Objective: The main objective of this study is to develop and assess an educational tool aimed at educating consumers about detecting health misinformation on WhatsApp. The secondary objective is to assess the association between demographic factors and knowledge levels.

Methods: The study used a single-arm, pre-post intervention design to evaluate the effectiveness of an educational video in improving participants’ ability to detect health-related misinformation in WhatsApp messages. In the first phase, an educational video intervention was developed and validated. In the second phase, participants were invited to complete a web-based survey that consisted of pre-evaluation questions, followed by the educational video intervention. Subsequently, they were asked to answer the same questions as the postevaluation questions.

Results: The web-based survey received 485 responses. The completion rate was 99.6% (n=483). Statistically significant associations existed between knowledge level and age, gender, employment, and region of residence (P<.05). The video intervention did elicit a statistically significant change in the participants’ abilities to identify misinformation in WhatsApp messages (z=−6.887; P<.001). Viewing the video was associated with increased knowledge about the following concepts: checking the “forwarded” label (P<.001), looking for spelling and grammatical errors (P<.001), analyzing the facts (P=.03), checking links (P=.002, P=.001), and assessing the photos and videos (P<.001). There was a statistically significant difference in knowledge level before and after the intervention (P<.001).

Conclusions: This study developed and evaluated the effectiveness of an educational video intervention to improve health misinformation identification on WhatsApp among the Saudi Arabian population. The results indicate that educational videos can be valuable tools for improving participants’ abilities to identify misinformation. The outcomes of this research can contribute to our understanding of what constitutes an effective tool for enhancing health misinformation awareness. Such interventions...
may be particularly useful in combating misinformation among Arabic-speaking populations on WhatsApp, which may ultimately improve eHealth literacy. Limiting the prevalence and impact of misinformation allows people to make better-informed health decisions.

(MJMIR Form Res 2024;8:e50211) doi: 10.2196/50211

KEYWORDS
misinformation; education; WhatsApp; intervention; pre-postintervention design; health literacy; educational; video; videos; consumer; consumers; patient education; survey; surveys; web-based information; health information; reliability; accuracy; reliable; social media

Introduction

Background

Researching health problems and learning about health via the internet has become a prevalent practice [1]. The level of credibility of this health-related information and the way it is used by patients, caregivers, and other health consumers has garnered the attention of health care providers and authorities [2]. There are many inaccurate sources of information on the internet, and this can lead to users becoming misinformed. According to Chou et al [3], health misinformation is defined as any health-related factual claim that is false according to recent scientific evidence. Misinformation about health can adversely affect quality of life and even increase one’s mortality risk [1].

When the COVID-19 pandemic first started, the amount of information related to this new global pandemic increased at an unprecedented rate. The volume of information, as well as the rate at which new information appeared, increased rapidly [4]. Global pandemics such as COVID-19 are likely to lead to the increased spread of misinformation as people explore massive amounts of information about the disease and its health implications. The term “infodemic” is used to describe the current media environment, which is characterized by an overflow of both true and false information. During the pandemic, individuals generally look for accurate, unbiased information, but these sources may be hidden among misinformation spread through the infodemic [5].

Due to its capacity to rapidly disseminate information, social media can serve as a platform for the propagation of misinformation. The abundance of available information can lead to the predominance of misinformation, thus negatively affecting cognitive, logical, and decision-making capacities. WhatsApp, Twitter, and Facebook are the most commonly used social media platforms for spreading false information. Since the beginning of the COVID-19 pandemic, internet use has expanded worldwide, which has resulted in the proliferation of incorrect information via social media [6].

Saudi Arabia, with a population of over 35 million, is the second largest Arab country [7]. The COVID-19 pandemic has had a significant impact on internet usage in Saudi Arabia, with a reported increase to 91.2% (n=28,775,889) in 2020. This represents a rise of 2.6 percentage points compared to the previous year [8]. Alshareef and Alotiby [6] used a web-based survey to investigate the most widely used social media platforms in Saudi Arabia, the proportions of Saudi Arabians who used these platforms to share information, and these users’ perceptions of the medical information shared on these platforms. According to their survey results, WhatsApp was used by 52.4% (n=144) of health care workers and 51.3% (n=500) of non–health care workers to circulate information.

The findings of their study concluded that WhatsApp is the most commonly used social network among Saudi Arabians. COVID-19–related information is, therefore, more likely to be shared on this application [6].

Another study by Alasmari et al [9] found that social media platforms, with their capacity to quickly disseminate information, comprised the primary source of falsehoods spread in the community. Based on an examination of the social media platforms, the study revealed that WhatsApp users accounted for approximately 46% (n=41) of rumor sources on the internet in Saudi Arabia.

Additionally, research by Tan et al [10] examined daily WhatsApp use for receiving, forwarding, or discussing COVID-19–related content a in 1-week period. The results indicate that almost every respondent participated in conversations about COVID-19. However, users were more likely to share or receive forwarded messages than to engage in active, original conversations about COVID-19. A high volume of forwarded messages was observed; this is concerning because the developers of WhatsApp have linked forwarded messages with misinformation.

People rarely assess the content of messages before sharing them on social media platforms, and they frequently fail to verify whether the messages are accurate. Educating consumers about identifying misinformation and dealing with the infodemic is essential. The false information epidemic compromises public health as misinformation spreads throughout social media. It is critical to increase awareness about the nature of social media and how to use it effectively. Personal responsibility is the first and most crucial step in safeguarding our community from the harmful phenomena of misinformation [11].

To effectively access health-related information on the internet, consumers must be able to assess the quality of the information that they find. This is a crucial aspect of eHealth literacy. It remains difficult for digital health consumers to determine the quality of the information placed in front of them. The problem becomes more complex as the digital information environment becomes more complicated and heterogeneous, especially with the rise of social media, where anyone can spread information about health and where low-quality and misleading information...
spreads rapidly. Interventions are urgently needed to address this public health problem [12].

Several interventions have been developed to assist consumers in verifying the validity of digital health information [12]. A systematic review by Cusack et al [13] examined studies on educational interventions that aimed to improve knowledge of essential concepts, enabling health interventions to be evaluated for their impacts. According to the study, educational interventions, at least in the short term, can increase people's knowledge and skills in evaluating health claims.

For the detection of health misinformation, interventions have been established based on instruments that allow anyone, including those with no prior medical background, to differentiate fact from fiction. However, these tools were designed for lengthy texts (such as text found on websites) and have not yet been validated for detecting health misinformation in social media content [14]. The majority of the tools developed were used to assess the quality of websites that provided health information. Considering the increased use of social media platforms—primarily WhatsApp—in Saudi Arabia for sharing health information, it is essential to develop tools that help consumers assess the credibility of messages and detect misinformation.

The first World Health Organization Infodemiology conference for managing the infodemic suggested evidence-based analysis and interventions to reduce the harmful effects of health misinformation during acute health events. Among the recommendations was the development of interventions that address factors that impact trust and resilience to misinformation at the individual, community, cultural, and societal levels [15].

Theoretical Background

A low level of health literacy has been recognized as one of the factors contributing to the infodemic. Other contributing factors include the widespread use of social media, quick publication processes, and preprint services. Rumor-spreading behavior also plays a role in the infodemic as do anxiety, distress, and fear [16]. According to a systematic review by Diviani et al [17], health literacy is essential when evaluating digital health information. Individuals’ abilities to find, evaluate, and use health information empowers them to actively deal with the misinformation they encounter on social media. In order to prevent people from automatically accepting health rumors as facts, health literacy must be improved [18].

In this research, the educational intervention concept was guided by the inoculation theory and the message interpretation process theory. According to the inoculation theory, previous experience helps individuals combat future attacks [19]. For example, literacy interventions may help audiences resist harmful media messages by providing them with the knowledge and skills necessary to reject them [20]. Based on the message interpretation process theory, exposure to message interventions influences subsequent decision-making when dealing with harmful information [21]. Both theories identify the role of an intervention or prior messages in influencing the cause of action [22].

The choice of the intervention media was guided by the cognitive theory of multimedia learning, which is built from the cognitive load theory and states that working memory contains 2 channels for acquiring and processing information: an auditory or verbal channel and a visual or pictorial channel. Although each channel has a limited capacity, the 2 can be used together to integrate new information more easily. Working memory can function at its best when both channels are used. However, 1 or both channels can become overloaded by a heavy cognitive load. It is possible to improve learning through the use of multimedia learning materials that manage the cognitive load across both channels. Furthermore, the cognitive theory of multimedia learning states that any learning should involve cognitive processing to be meaningful. Cognitive processing requires a learner to pay attention to the material presented, organize it mentally, and integrate it into prior knowledge [23].

Objectives

The main objective of this study is to develop and assess an educational tool aimed at educating consumers about detecting health misinformation on WhatsApp. The secondary objective is to assess the association between demographic factors and knowledge levels.

Methods

Study Design

The study used a single-arm, pre-post intervention design to evaluate the effectiveness of educational video in improving participants’ ability to detect health-related misinformation in WhatsApp messages. The study’s first phase was developing and validating an educational video intervention. In the second phase, participants were invited to complete a web-based survey that contained pre-evaluation questions, the intervention, and postevaluation questions.

Participants

A web-based survey was distributed among the general Saudi population from November 24 to December 25, 2022. The survey was disseminated through social media networks (WhatsApp, Instagram, Twitter, Facebook, and Telegram), and the data were collected using Google Forms.

It has been estimated that 82% (n=29.50 million) of Saudis use social networks daily, with varying usage rates among different platforms. Among these platforms, WhatsApp is the most widely used social network with 87.4% (n=30.67 million) of internet users in Saudi Arabia, followed by Instagram (n=27.40 million, 78.1%), Twitter (n=25.23 million, 71.9%), Facebook (n=22.25 million, 63.4%), and Telegram (n=20.88 million, 59.5%) [24]. In order to target a wide range of the population, the web-based survey was disseminated across all of these social media platforms.

The study population consisted of social media users in the general population of the Kingdom of Saudi Arabia. The inclusion criteria were (1) having the ability to complete an anonymous survey questionnaire on the internet, (2) being at least 18 years of age, and (3) understanding Arabic.
The sample size was calculated using the Raosoft sample size calculator, based on the total population of Saudi Arabia (n=35,013,414), with a 95% CI [7.25]. This calculation yielded a minimum sample size of 385 using absolute error or precision of 0.05. This sample size is sufficient to detect a difference between pre and postscore with an effect size of 0.15 (small effect size) using a power of 80% and \( \alpha \) of .05.

The study population was targeted using a convenience sampling technique with no predetermined sampling frame. Convenience sampling is a nonprobability method in which individuals are sampled simply because they are “convenient” data sources [26].

Several specific methods were used for the recruitment process. First, as there are social accounts run by the public to share news and announcements related to each region in Saudi Arabia, the survey was distributed to these public social networking groups on various social media platforms. Second, the researchers approached social media influencers on different platforms to spread the survey to more participants. Third, the researchers asked all their social media contacts to consider completing the survey and sharing it with their contacts on social networks.

The web-based survey had 4 sections. In the first section, the participants were asked to provide demographic information. In the second section, they were given a set of pretest questions asking them to identify whether a WhatsApp message contained correct or false information. In the third section, the participants were shown an educational video. After finishing the video, the participants moved to the last section, which contained the same set of questions as the pretest.

There were no records of participant identity, and confidentiality was ensured. Upon completion of the survey, a message of thanks appeared. No incentives were offered for completing the survey.

**Intervention (Educational Video)**

**Educational Video Design**

This study used a short video intervention. The content of the educational video was developed based on three sources: (1) the recommendations on WhatsApp’s official website regarding how to prevent the spread of misinformation; (2) the World Health Organization’s (WHO’s) advice on how to navigate the infodemic and identify misinformation; and (3) the CRAAP test, a tool for evaluating the quality of a social media source by assessing its currency, relevance, authority, accuracy, and purpose [27-29]. The educational tool introduced 6 concepts that could be used to assess and identify misinformation in WhatsApp messages. These concepts included checking the “forwarded” label, looking for spelling and grammatical errors, reading beyond the headline, analyzing the facts, checking links, and assessing photos and videos (Table 1).

### Table 1. Concepts used in the video to evaluate and identify misinformation in WhatsApp messages.

<table>
<thead>
<tr>
<th>Concept</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Check the “forwarded” label</td>
<td>WhatsApp [27]</td>
</tr>
<tr>
<td>Look for spelling and grammatical errors</td>
<td>WHOa, WhatsApp, and CRAAP Test [27-29]</td>
</tr>
<tr>
<td>Read beyond the headline</td>
<td>WHO [28]</td>
</tr>
<tr>
<td>Analyze the facts</td>
<td>WHO, WhatsApp, and CRAAP Test [27-29]</td>
</tr>
<tr>
<td>Check links</td>
<td>CRAAP Test [29]</td>
</tr>
<tr>
<td>Assess the photos and videos</td>
<td>WHO [28]</td>
</tr>
</tbody>
</table>

*aWHO: World Health Organization.*

The design of the educational video was based on literature guidelines for the design of health education messages [30]. Following Hugo recommendation, the construction of the educational material included the consideration of communication principles and sociocultural factors, including the literacy levels and language preferences of the audience, to design appropriate messages. When designing the audiovisual content, simplicity (text and visual composition) and the audience’s emotional involvement were considered [30]. The educational video was developed in classical Arabic to make it accessible to a wider audience. Multimedia Appendix 1 shows the developed educational video, while screenshots of the educational video are shown in Figure 1.
Educatioonal Video Validation Process

Yusoff [31] recommended that when validating the content of a tool, a minimum of 6 (but no more than 10) experts should be involved in the assessment process. The validation of the educational material included assessments by 7 experts (Multimedia Appendix 2). For this study, it was essential that the specialists were well-versed in the Arabic language (spoken and written).

To identify the broad range of expertise needed, roles were categorized into different fields. Four roles were identified: (1) health informatics experts, (2) health education specialists, (3) infodemic managers, and (4) public health experts.

Each category was populated with a representative through purposeful sampling. A panel of experts was formed with at least 1 representative from each role; email was used to contact the representatives and provide information about the study. The validation forms and the educational video were delivered to the specialists via email.

The validation form was created by combining 2 tools from which items relevant to the study were selected. Questions Q2 through Q11 were adopted from the Educational Content Validation Instrument in Health developed by Leite et al [32], and questions Q1 and Q12 through Q17 were adopted from an audiovisual content evaluation instrument constructed by Rosa et al [33].

The final evaluation form contained 17 questions covering three areas: objectives, structure and presentation, and audiovisuals (Multimedia Appendix 3). The “objectives” section focused on purposes and goals, whereas the “structure and presentation” section emphasized organization, structure, strategy, sufficiency, and consistency. As for the “audiovisual” area, the emphasis was on the technological aspect. A score of 0 indicated disagreement, 1 indicated partial agreement, and 2 indicated strong agreement with the value of the items [34].

Educatioonal Video Validation Result

A content validity index was used to analyze the results. Content validity indexes can be computed in 2 ways. One type of validity is item-level content validity indexes (I-CVIs), which consider the content validity of individual items. The other type is scale-level content validity indexes, which involve a scale’s overall content validity [35]. For the scale-level content validity,
calculations were conducted using the scale-level content validity index averaging method (S-CVI/Ave) as recommended by Polit and Beck [36].

The calculations were carried out manually. The items ranked “disagree” were scored as 0, whereas the items ranked “partially agree” and “strongly agree” were scored as 1 [34].

To obtain excellent content validity, the content of educational videos must have items with I-CVIs above 0.78 for (6 to 10 experts) and an S-CVI/Ave of 0.90 or higher [36]. When the I-CVI is below 0.78 and the S-CVI/Ave is below 0.90, the content modification should be considered for that particular educational video area.

As shown in Table 2, all items had I-CVIs greater than 0.78 (78%), indicating agreement between the experts’ answers. In terms of scale evaluation, all the 3 areas (objectives, structure and presentation, and audiovisual) had S-CVI/Aves above 0.90 (90%). In the objectives area, item 5 (“stimulates interest in the theme”) had the lowest specialist agreement score (6 out of 7 or an I-CVI of 0.86). The overall S-CVI/Ave for the objectives area was 0.97.

<table>
<thead>
<tr>
<th>Item</th>
<th>Expert 1</th>
<th>Expert 2</th>
<th>Expert 3</th>
<th>Expert 4</th>
<th>Expert 5</th>
<th>Expert 6</th>
<th>Expert 7</th>
<th>Experts In Agreement</th>
<th>I-CVIa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>0.86</td>
</tr>
<tr>
<td>Q2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>0.86</td>
</tr>
<tr>
<td>Q3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>0.86</td>
</tr>
<tr>
<td>Q4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>0.86</td>
</tr>
<tr>
<td>Q5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>0.86</td>
</tr>
<tr>
<td>Proportion relevance</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>0.86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item</th>
<th>Expert 1</th>
<th>Expert 2</th>
<th>Expert 3</th>
<th>Expert 4</th>
<th>Expert 5</th>
<th>Expert 6</th>
<th>Expert 7</th>
<th>Experts In Agreement</th>
<th>I-CVIa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>0.86</td>
</tr>
<tr>
<td>Q7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>0.86</td>
</tr>
<tr>
<td>Q8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>0.86</td>
</tr>
<tr>
<td>Q9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>0.86</td>
</tr>
<tr>
<td>Q10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>0.86</td>
</tr>
<tr>
<td>Q11</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>0.86</td>
</tr>
<tr>
<td>Q12</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>0.86</td>
</tr>
<tr>
<td>Proportion relevance</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item</th>
<th>Expert 1</th>
<th>Expert 2</th>
<th>Expert 3</th>
<th>Expert 4</th>
<th>Expert 5</th>
<th>Expert 6</th>
<th>Expert 7</th>
<th>Experts In Agreement</th>
<th>I-CVIa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q13</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>0.86</td>
</tr>
<tr>
<td>Q14</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>0.86</td>
</tr>
<tr>
<td>Q15</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>0.86</td>
</tr>
<tr>
<td>Q16</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>0.86</td>
</tr>
<tr>
<td>Q17</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>0.86</td>
</tr>
<tr>
<td>Proportion relevance</td>
<td>1</td>
<td>1</td>
<td>0.8</td>
<td>0.8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

aI-CVI: item-level content validity index.

bThe S-CVI/Ave and average proportion of items judged as relevance across the 7 experts for objectives is 0.97.

cThe S-CVI/Ave and average proportion of items judged as relevance across the 7 experts for structure and presentation is 1.

dThe S-CVI/Ave and average proportion of items judged as relevance across the 7 experts for audiovisual is 0.94.

The structure and presentation area had the highest level of agreement (100% S-CVI/Ave). In the audiovisual area, 2 items (“the illustrations are expressive and sufficient” and “the characters/images are appropriate for the target audience”) had scores of 6 out of 7 or I-CVIs of 0.86. The overall S-CVI/Ave for the audiovisual area was 0.94.
Data Collection Tool

Development of the Survey

The study aimed to develop and assess an interventional tool to educate the Saudi population on how to identify health misinformation in WhatsApp messages. The developed Arabic survey contained 4 sections: demographic data, pretest questions, educational intervention, and posttest questions.

The first section (sociodemographic information) included 7 questions about each participant’s background information, including gender, age, educational level, employment status, region of residence, city of residence, and nationality. The second section included 8 questions that assessed each participant’s ability to identify misinformation based on the WhatsApp messages evaluation concepts mentioned in Table 1. Multimedia Appendix 4 shows the complete list of questions that were assigned to the concepts. The third section included the educational tool, which discussed 6 concepts that could be used to identify misinformation in WhatsApp messages. The fourth section included the same 8 questions as the second domain to measure the effectiveness of the educational tool.

The assessment questions were based on real examples representing the different domains of WhatsApp messages. The 8 pre and posttest questions included 5 messages with misinformation (based on messages circulated during the pandemic) and 3 with correct information (obtained from the Saudi Ministry of Health’s official website) [37]. The evaluation concepts, selection of the messages, and their relevancy were assessed by 2 authors (EA and SA). EA collected the messages and placed them under each domain, and SA assessed the relevancy; any uncertainty was resolved by consensus. Finally, a summary of the study’s goal was included in the survey, as was a statement assuring the respondents’ confidentiality.

The highest possible score for the survey questions was 8 points (correct answers were scored as 1 point; incorrect answers and responses of “I do not know” were scored as 0 points). We used a modified Bloom cutoff value of 75% to categorize participants’ knowledge. As a result, we considered participants with scores ≥75% to have high knowledge. Participants with scores below 75% were considered to have low knowledge. This cutoff value is based on previous publications [38,39].

Pilot Survey

The web-based survey was pilot tested, and a total of 31 participants responded. During the first piloting stage, the survey was sent to 15 participants, 3 of whom commented that the instructions needed clarification. Subsequently, an instruction section containing a description of the other survey sections was added at the beginning, and the survey was distributed again to 16 participants. In the second stage, no further comments were received; all participants indicated that the survey was clear. The internal consistency of the final survey was measured using Cronbach α. The scale had a Cronbach α of .847, demonstrating good internal reliability. The final version of the survey can be found in Multimedia Appendix 5.

Ethical Considerations

The study was approved by King Abdullah International Medical Research Center (reference RYD-22-419812-107000). The survey included a summary of the study’s purpose and a statement that, by completing the survey, the respondents agreed to participate in this research. The confidentiality of the study participants was ensured by not collecting identifiable data, encrypting files, and requiring a password to open or modify files.

Statistical Analysis

The demographic characteristics of the participants were reported using descriptive statistics, such as frequencies and percentages. Categorical variables were analyzed using chi-square tests to determine the associations between the demographic variables and the knowledge levels of the participants. A modified Bloom cutoff was used to categorize the knowledge levels. The normality of the variables was analyzed using the Kolmogorov-Smirnov test and the Shapiro-Wilk test. The median scores from before and after the educational video were compared using the Wilcoxon signed-rank test. The McNemar test for categorical data was used to compare the answers and differences in knowledge levels before and after the educational video intervention. The analyses were performed using SPSS Statistics (version 29.0; IBM). A P-value of .05 or less was considered statistically significant.

Results

Characteristics of Study Participants

The web-based survey received 485 responses, and 2 did not agree to participate, giving a 99.6% (n=483) completion rate. In total, 483 responses were analyzed. The socioeconomic characteristics of the participants are shown in Table 3. Most of the study respondents (n=457, 94.6%) were Saudis. More than half of the participants (n=300, 62.1%) were female, and more than half were in the age range of either 18-24 or 25-34 years (n=130, 26.9% and n=173, 35.8%, respectively). More than half of the sample (n=275, 56.9%) had bachelor degrees. With regard to the employment status, 45.5% (n=220) of the respondents were employed. The highest number of participants came from the eastern region (n=181, 37.5%), followed by the central (n=132, 27.3%) and western (n=83, 17.2%) regions.
Table 3. Participant demographics of the full sample (N=483) who participated in a web-based survey about the ability to identify health misinformation on WhatsApp messages between November and December 2022.

<table>
<thead>
<tr>
<th></th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nationality</strong></td>
<td></td>
</tr>
<tr>
<td>Saudi</td>
<td>457 (94.6)</td>
</tr>
<tr>
<td>Non-Saudi</td>
<td>26 (5.4)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>130 (26.9)</td>
</tr>
<tr>
<td>25-34</td>
<td>173 (35.8)</td>
</tr>
<tr>
<td>35-44</td>
<td>99 (20.5)</td>
</tr>
<tr>
<td>45-54</td>
<td>47 (9.7)</td>
</tr>
<tr>
<td>55 or older</td>
<td>34 (7)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>183 (37.9)</td>
</tr>
<tr>
<td>Female</td>
<td>300 (62.1)</td>
</tr>
<tr>
<td><strong>Educational level</strong></td>
<td></td>
</tr>
<tr>
<td>High school and below</td>
<td>94 (19.4)</td>
</tr>
<tr>
<td>Diploma</td>
<td>39 (8.1)</td>
</tr>
<tr>
<td>Bachelor degree</td>
<td>275 (56.9)</td>
</tr>
<tr>
<td>Postgraduate degree</td>
<td>75 (15.5)</td>
</tr>
<tr>
<td><strong>Employment status</strong></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>114 (23.6)</td>
</tr>
<tr>
<td>Employed</td>
<td>220 (45.5)</td>
</tr>
<tr>
<td>Not employed</td>
<td>122 (25.3)</td>
</tr>
<tr>
<td>Retired</td>
<td>27 (5.6)</td>
</tr>
<tr>
<td><strong>Region of residence</strong></td>
<td></td>
</tr>
<tr>
<td>East</td>
<td>181 (37.5)</td>
</tr>
<tr>
<td>Central</td>
<td>132 (27.3)</td>
</tr>
<tr>
<td>West</td>
<td>83 (17.2)</td>
</tr>
<tr>
<td>North</td>
<td>48 (9.9)</td>
</tr>
<tr>
<td>South</td>
<td>39 (8.1)</td>
</tr>
</tbody>
</table>

Despite some variations, the sample matched the age and sex distribution of the Saudi Arabian population. Similarities in regional distribution between our sample and the populations of certain regions were also evident in the sample, with 27.3% (n=132) from the central region and 9.9% (n=48) from the northern region aligning with the national census in Saudi Arabia (n=5,365,700, 28.5% and n=1,877,108, 9.9%, respectively). In the eastern region, our sample showed a higher representation at 37.5% (n=181) compared with 15.7% (n=2,949,854) reported in the census data [40]. Our sample also showed a difference in educational level, with 56.9% (n=275) of participants holding bachelor degrees, in contrast to the national statistic of 23% (n=2,812,477) [41].

**Association Between Knowledge Level and Demographic Variables**

The highest possible score for the survey questions was 8 points (correct answers were scored as 1 point; incorrect answers and responses of “I do not know” were scored as 0 points). We used a modified Bloom cutoff value of 75% (6 points) to categorize the participants’ knowledge. A knowledge score of ≥6 indicated a high level of knowledge, while a score of <6 indicated a low level of knowledge.

The associations between knowledge about identifying misinformation in WhatsApp messages and demographic variables were assessed using chi-square tests (Table 4). There were statistically significant associations between knowledge level and age, sex, employment, and region of residence (P<.05).
Table 4. Chi-square tests to examine the association between knowledge about identifying misinformation in WhatsApp messages before the intervention and demographic variables (N=483).

<table>
<thead>
<tr>
<th>Factor</th>
<th>Low knowledge, n (%)</th>
<th>High knowledge, n (%)</th>
<th>Chi-square (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nationality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saudi</td>
<td>136 (29.8)</td>
<td>321 (70.2)</td>
<td>0.529 (1)</td>
<td>.47</td>
</tr>
<tr>
<td>Non-Saudi</td>
<td>6 (23.1)</td>
<td>20 (76.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>18-24</td>
<td>22 (16.9)</td>
<td>108 (83.1)</td>
<td>43.030 (4)</td>
<td></td>
</tr>
<tr>
<td>25-34</td>
<td>37 (21.4)</td>
<td>136 (78.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-44</td>
<td>41 (41.4)</td>
<td>58 (58.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-54</td>
<td>22 (46.8)</td>
<td>25 (53.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55 and above</td>
<td>20 (58.8)</td>
<td>14 (41.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Male</td>
<td>64 (35)</td>
<td>119 (65)</td>
<td>4.409 (1)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>78 (26)</td>
<td>222 (74)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td>.06</td>
</tr>
<tr>
<td>High school and below</td>
<td>35 (37.2)</td>
<td>59 (62.8)</td>
<td>7.289 (3)</td>
<td></td>
</tr>
<tr>
<td>Diploma</td>
<td>16 (41)</td>
<td>23 (59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor degree</td>
<td>71 (25.8)</td>
<td>204 (74.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postgraduate degree</td>
<td>20 (26.7)</td>
<td>55 (73.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Employment</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Student</td>
<td>19 (16.7)</td>
<td>95 (83.3)</td>
<td>21.378 (3)</td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>71 (32.3)</td>
<td>149 (67.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not employed</td>
<td>36 (29.5)</td>
<td>86 (70.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retired</td>
<td>16 (59.3)</td>
<td>11 (40.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Region of residence</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>East</td>
<td>61 (33.7)</td>
<td>120 (66.3)</td>
<td>35.330 (4)</td>
<td></td>
</tr>
<tr>
<td>Central</td>
<td>25 (18.9)</td>
<td>107 (81.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>13 (15.7)</td>
<td>70 (84.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North</td>
<td>27 (56.3)</td>
<td>21 (43.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South</td>
<td>16 (41)</td>
<td>23 (59)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Effectiveness of the Intervention**

The Kolmogorov-Smirnov test and the Shapiro-Wilk test indicated that the knowledge scores before and after the educational video were not normally distributed (P<.001). Since the distribution was not symmetric, it is negatively skewed, and there are a few outliers on the left side contributing to the skewness; nonparametric tests were used to assess statistical significance.

The median scores were assessed both before (median = 7, IQR = 5-8) and after (median = 8, IQR = 6-8) the video intervention, using the Wilcoxon signed-rank test. This comparison revealed that the video intervention did elicit a statistically significant change in the participants’ abilities to identify misinformation in WhatsApp messages (z=-6.887; P<.001).

**Knowledge Questions**

The proportions of correct answers per individual test question before and after the video intervention were compared using the McNemar test. Significant differences in the participants’ pre- and postintervention knowledge about identifying misinformation were found for specific questions (Table 5). Viewing the video was associated with increased knowledge about the following concepts: checking the “forwarded” label (P<.001), looking for spelling and grammatical errors (P<.001), analyzing the facts (P=.03), checking links (P=.002, P=.001), and assessing the photos and videos (P<.001).
Table 5. Participants’ answers before and after viewing the educational video (N=483).

<table>
<thead>
<tr>
<th>Domains and answer</th>
<th>Preintervention, n (%)</th>
<th>Postintervention, n (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Check the “Forwarded” label</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Correct</td>
<td>387 (80.1)</td>
<td>423 (87.6)</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>96 (19.9)</td>
<td>60 (12.4)</td>
<td></td>
</tr>
<tr>
<td>Look for spelling and grammatical errors</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Q2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>357 (73.9)</td>
<td>402 (83.2)</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>126 (26.1)</td>
<td>81 (16.8)</td>
<td></td>
</tr>
<tr>
<td>Read beyond the headline</td>
<td></td>
<td></td>
<td>.061</td>
</tr>
<tr>
<td>Q4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>416 (86.1)</td>
<td>430 (89.0)</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>67 (13.9)</td>
<td>53 (11)</td>
<td></td>
</tr>
<tr>
<td>Analyze the facts</td>
<td></td>
<td></td>
<td>.028</td>
</tr>
<tr>
<td>Q5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>418 (86.5)</td>
<td>435 (90.1)</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>65 (13.5)</td>
<td>48 (9.9)</td>
<td></td>
</tr>
<tr>
<td>Check links</td>
<td></td>
<td></td>
<td>.54</td>
</tr>
<tr>
<td>Q3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>387 (80.1)</td>
<td>402 (83.2)</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>96 (19.9)</td>
<td>81 (16.8)</td>
<td></td>
</tr>
<tr>
<td>Q6</td>
<td></td>
<td></td>
<td>.002</td>
</tr>
<tr>
<td>Correct</td>
<td>326 (67.5)</td>
<td>354 (73.3)</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>157 (32.5)</td>
<td>129 (26.7)</td>
<td></td>
</tr>
<tr>
<td>Q8</td>
<td></td>
<td></td>
<td>.001</td>
</tr>
<tr>
<td>Correct</td>
<td>377 (78.1)</td>
<td>403 (83.4)</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>106 (21.9)</td>
<td>80 (16.6)</td>
<td></td>
</tr>
<tr>
<td>Assess the photos and videos</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Q7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>333 (68.9)</td>
<td>368 (76.2)</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>150 (31.1)</td>
<td>115 (23.8)</td>
<td></td>
</tr>
</tbody>
</table>

Improvement in Knowledge Level

The health misinformation education intervention involved 483 participants. Pretest results showed that 70.6% (n=341) of participants had high knowledge (score ≥6), while 29.4% (n=142) had low knowledge (score>6). After the posttest, 10.6% (n=51) of the sample had improved to high knowledge and 3.3% (n=16) had lower scores, indicating 77.8% (n=376) had a score of 6 or above. McNemar test determined that there was a statistically significant difference in knowledge level before and after the intervention (P<.001; Table 6).

Table 6. McNemar test to compare knowledge level before and after the intervention (n=483; P<.001).

<table>
<thead>
<tr>
<th>Before, n (%)</th>
<th>After, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Low</td>
<td>91 (18.9)</td>
</tr>
<tr>
<td>High</td>
<td>16 (3.3)</td>
</tr>
</tbody>
</table>
Discussion

Overview
This study aimed to design and evaluate the effectiveness of an educational video to improve the abilities of participants in Saudi Arabia to identify health misinformation within the WhatsApp app. The study used a single-arm, pre-post intervention design and was conducted on the web. The effectiveness of the intervention was assessed. Furthermore, the participants’ knowledge levels about identifying misinformation were assessed before and after the intervention as were the associations between the participants’ characteristics and their knowledge.

The proliferation of health-related misinformation on social media has raised public health concerns in many countries [42]. Chen et al [43] found that people with limited health literacy were more likely to trust health-related information found on social media and blogs. Thus, improving the public’s ability to evaluate health information may be necessary. In Saudi Arabia, WhatsApp is the most popular social network, which is used by 87.4% (n=30.67 million) of internet users [6,24]. This platform has been identified as one on which misinformation may be easily spread [44]; therefore, it was the platform on which this study focused.

Principal Results
The participants’ ability to identify misinformation in WhatsApp messages significantly improved following the educational intervention (P<.001). This result supports the finding of a systematic review by Cusack et al [13], which showed that, at least in the short term, educational interventions could improve knowledge and skills. This finding is also in line with the message interpretation process theory and the inoculation theory, in which interventions and prior messages are identified as factors that effectively protect against the harm caused by misinformation [19,21].

Additionally, the findings of this study suggest that literacy interventions combined with visual multimedia may improve misinformation detection. Aputre et al [45] found that participants who received visual multimedia education had better knowledge of literacy concepts than those who were educated without visual multimedia. Thus, as previously mentioned in this study, multimedia enhances memory, as stated in the cognitive theory of multimedia learning. The receiver’s exposure to the various message components makes it easier to integrate new knowledge [23].

In this study, the items that assessed the following concepts (checking the “forwarded” label, looking for spelling and grammatical errors, analyzing the facts, assessing the photos and videos, and checking links) were significantly associated with improvements in the participants’ knowledge (P<.05 for all). However, the item related to the concept of reading beyond the headline was not significantly associated with improvement (P>.05).

It was noticed that WhatsApp messages with misinformation are characterized by requests to forward the message to many people. Further, many forwarded messages include fake information sources such as links or names and are vague about timelines, authors, and origins. Consequently, most forwarded messages are found to contain misinformation [46].

By recognizing the most prominent characteristics of health misinformation, users can improve their abilities to identify it on social media. Some studies have proposed criteria such as accuracy, authority, objectivity, and currency, but it is challenging for laypeople to evaluate these indicators. The role of such criteria is quite limited for general users, who, by definition, are not professionals [47]. Li et al [47] proposed a feature scheme and incorporated semantic, grammatical, and peripheral features of messages in evaluating their credibility. Their developed feature scheme allowed users to improve both their abilities to recognize health misinformation and their levels of digital health literacy.

This study found a significant association between knowledge level and age, sex, employment, and region of residence (P<.05 for all). Bapaye and Bapaye [48] noted that those engaged in elementary occupations and those older than 65 years of age were most likely to get false information from WhatsApp in a developing country. Workers in the health care industry were not immune from the impact of false information and were found to be just as susceptible as those in other professions.

The pretest results showed that 70.6% (n=341) of participants had good levels of knowledge about identifying misinformation on WhatsApp. There may have been factors contributing to an increase in knowledge, such as public awareness campaigns and government efforts. During the pandemic, the Saudi Arabian Ministry of Health conducted a comprehensive media campaign that included television, websites, and social media. Taking advantage of social media platforms, the Ministry of Health also engages with the public and the media. In addition to these early initiatives, efforts have been made to combat rumors and misinformation and engage the public in prevention and control measures [49].

Higher health literacy levels are associated with more favorable perceptions of health information. However, health literacy varies depending on the situation, and thus even those with high levels of health literacy may need help occasionally. For instance, those unfamiliar with medical language may find it challenging to distinguish between materials that provide accurate information and those with inaccurate information. Health care professionals and organizations must evaluate the population’s level of health literacy in order to ensure that people have access to adequate information when it matters most. Strategies like awareness-raising campaigns, community engagement, educational interventions, and training programs should be implemented when needed [50]. In accordance with the first Infodemiology Conference of the World Health Organization, public health authorities must create, evaluate, implement, and adapt tools and strategies for managing infodemics in acute public health crises in a manner that is suitable for their countries and situations [15]. This study’s findings may provide insight to public health authorities about developing an appropriate intervention for the population.
Strengths and Limitations
The strength of this study is that it involved developing and validating an educational video in the Arabic language to identify misinformation on WhatsApp. Few Arabic educational materials exist to combat misinformation. The study’s limitations include some sampling bias due to the use of convenience sampling, which is a nonprobability sampling technique. While this technique may limit the generalizability of the results, it was appropriate for our study because it is more cost-effective, faster, and more direct than other sampling techniques.

Conclusions
Health misinformation is an issue threatening public health because it dominates social media. Training people on the characteristics and practical applications of social media is urgently necessary. This study developed and evaluated the effectiveness of an educational video intervention to improve health misinformation identification on WhatsApp among the Saudi Arabian population. The results indicate that educational videos can be valuable tools for improving participants’ abilities to identify misinformation. The outcomes of this research can contribute to our understanding of effective tools for enhancing health misinformation awareness. These interventions can be particularly useful in combating misinformation in Arabic-speaking populations on WhatsApp, which may ultimately improve eHealth literacy. Limiting the prevalence and impact of misinformation allows people to make better-informed health care decisions. Our findings may also be helpful for health care professionals and organizations deciding on interventions suitable for providing access to adequate information to certain populations when needed.

Acknowledgments
The authors appreciate the support provided by the Office of Research in KSAU-HS, especially Mr Ahmed Aldakhil, for his valuable biostatistics consultations. The authors gratefully acknowledge King Abdullah International Medical Research Center for funding the publication of this paper.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions
The study was designed and planned by EA and SA. EA took on the task of developing the educational video and survey. Data collection, results interpretation, and manuscript writing were all carried out by EA, SA reviewed the work. EA and SA discussed the findings, contributed to the final manuscript, and shaped the research.

Conflicts of Interest
None declared.

Multimedia Appendix 1
The developed educational video (Health Misinformation Identification on WhatsApp).
[MP4 File (MP4 Video), 11898 KB - formative_v8i1e50211_app1.mp4 ]

Multimedia Appendix 2
Experts names and positions.
[DOCX File, 14 KB - formative_v8i1e50211_app2.docx ]

Multimedia Appendix 3
The video content validation form.
[DOCX File, 15 KB - formative_v8i1e50211_app3.docx ]

Multimedia Appendix 4
WhatsApp messages evaluation concepts and the corresponding survey questions used to assess them.
[DOCX File, 1738 KB - formative_v8i1e50211_app4.docx ]

Multimedia Appendix 5
The final version of health misinformation identification survey.
[DOCX File, 972 KB - formative_v8i1e50211_app5.docx ]

References


46. Durul SS. (Mis)information in baby boomers' WhatsApp messages. Eskişehir, Turkey: Anadolu University; 2020 Presented at: 17th International Symposium Communication in the Millenium; November 5-6, 2020; Istanbul, Turkey p. 290-298 URL: http://tinyurl.com/37ce63ax
Abbreviations

**I-CVI**: item-level content validity index

**S-CVI/Ave**: scale-level content validity index averaging method

**WHO**: World Health Organization
A Blended Intervention Targeting Emotion Dysregulation in Adults With Attention-Deficit/Hyperactivity Disorder: Development and Feasibility Study

Emilie S Nordby1,2, MSc; Frode Guribye3, PhD; Viktor Schønning4, MSc; Sander Lindholm Andersen1, MSc; Jonna Kuntsi5, PhD; Astrid J Lundervold2, PhD

1Division of Psychiatry, Haukeland University Hospital, Bergen, Norway
2Department of Biological and Medical Psychology, Faculty of Psychology, University of Bergen, Bergen, Norway
3Department of Information Science and Media Studies, Faculty of Social Sciences, University of Bergen, Bergen, Norway
4Department of Child and Adolescent Psychiatry, Division of Psychiatry, Haukeland University Hospital, Bergen, Norway
5Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London, United Kingdom

Abstract

Background: Many adults with attention-deficit/hyperactivity disorder (ADHD) experience difficulties related to emotion regulation. Such difficulties are known to substantially impact quality of life and overall functioning. Yet, there is a lack of treatment interventions specifically designed to address these challenges.

Objective: This study aimed to describe the development and assess the feasibility, along with the initial clinical outcomes, of a novel blended intervention for adults with ADHD. The blended intervention combines both face-to-face and digital components and is specifically designed to address emotion dysregulation in ADHD.

Methods: This intervention was an 8-week blended intervention combining weekly face-to-face group sessions with a supplementary digital companion app. The intervention is based on elements from dialectic behavioral therapy skills training and positive psychology. To evaluate its feasibility, we performed a 10-week feasibility study with an uncontrolled pre-post study design, including 16 adults with ADHD and co-occurring emotion dysregulation. The feasibility measures encompassed adherence, satisfaction, and perceived credibility of the intervention. Clinical outcomes were evaluated by self-reported symptoms of emotion dysregulation, inattention, hyperactivity-impulsivity, executive function, depression, anxiety, and a measure of quality of life. Paired sample 2-tailed t tests were used to analyze clinical outcomes with a Bonferroni-corrected significance level.

Results: Both treatment credibility and treatment satisfaction were rated favorably by the majority of the participants. In particular, the participants emphasized meeting others with ADHD as beneficial. In terms of adherence, 3 participants withdrew before initiating the intervention, while another 4 participants did not complete the intervention. On average, the participants who enrolled in the intervention attended 6.2 of the 8 group sessions and completed 6.7 of the 8 skills training modules in the companion app. In terms of clinical outcomes, there was a reduction in symptoms of emotion dysregulation from before to after the intervention ($d=2.0$). Significant improvements were also observed in measures of inattention ($d=1.1$) and hyperactivity-impulsivity ($d=0.9$). However, no significant improvements were found in the domains of depression, anxiety, quality of life, and executive functioning.

Conclusions: The results are encouraging, both in terms of feasibility and the preliminary clinical results on emotion dysregulation. The blended format, combining digital and face-to-face elements, may also seem to offer some advantages: the group-based format was valued as it facilitated peer interaction, while a rather high completion of modules in the companion app highlights its potential to enhance skills training between the group sessions. Future randomized controlled trials are called for to further evaluate the clinical effectiveness of the intervention.
Introduction

Attention-deficit/hyperactivity disorder (ADHD) affects approximately 2.6% of the adult population [1]. Although symptoms of ADHD initially present themselves in childhood, most individuals diagnosed with ADHD as children continue to experience symptoms and associated impairments into adulthood [2]. Moreover, there is a rising trend of individuals first receiving their ADHD diagnosis in adulthood [3]. ADHD manifests through symptoms of inattention, hyperactivity, and impulsivity that significantly disrupt the individual’s daily functioning [4]. Beyond these core symptoms, ADHD is accompanied by an array of co-occurring symptoms and difficulties, both somatic and psychological, which can further complicate and amplify the challenges associated with the diagnosis [5].

Emotion dysregulation is a common deficit observed across many mental health conditions [6]. It is characterized by challenges in effectively managing and modulating one’s emotions, including the emotional experience (eg, intensity and duration) and expression [7]. Emotion dysregulation commonly coexists with ADHD, impacting as many as 34%-70% of the adults with the diagnosis [8]. Emotion dysregulation also occurs in adults with ADHD without the presence of other comorbid diagnoses [9], giving support to the notion that it is a core component of ADHD [10-12]. The co-occurrence of emotion dysregulation with ADHD is linked to a range of negative outcomes, including occupational challenges, interpersonal conflicts, financial struggles, parenting stress, as well as tendencies for self-harm, illicit drug use, and suicidal ideation [13-15]. Moreover, research indicates that emotion dysregulation serves as an independent predictor of impairment in ADHD [16]. Its adverse influence on self-esteem and quality of life has further been observed to exceed the effects of inattention and hyperactivity [17]. Given the high prevalence of emotion dysregulation in ADHD and the associated negative outcomes, adults with ADHD should be offered treatment interventions that specifically aim to strengthen their emotion regulation skills.

While pharmacological treatment remains the main treatment approach for the management of ADHD in adults, the UK National Institute for Health and Care Excellence recommends supplementing with psychological interventions, such as psychoeducation or psychotherapy [18]. Psychological alternatives are especially critical in cases where the individual does not want to take medications or when medications either do not lead to sufficient clinical improvement or result in unwanted side effects. This may be particularly relevant for those with co-occurring emotion dysregulation, as ADHD medications appear to be less effective for these difficulties [19]. A systematic review and meta-analysis showed that traditional ADHD medications, including methylphenidate, atomoxetine, and lisdexamfetamine, only had small to moderate effects on emotion dysregulation among adults, which were significantly lower than the effect sizes reported for core symptoms of inattention, hyperactivity, and impulsivity [20]. As such, the authors of the abovementioned review emphasize that there is a need for more research on both pharmacological and psychological treatments targeting emotion dysregulation in adults with ADHD [20].

There is limited access to psychological treatments among adults with ADHD, and most available interventions mainly address the core symptoms of inattention and hyperactivity [21,22]. To the authors’ knowledge, there are currently 8 studies that have included emotion dysregulation as an outcome measure in studies of psychological interventions for adults with ADHD, including 2 randomized controlled trials [23-30]. However, only 2 of the studies reported emotion dysregulation to be the primary treatment target [24,26]. Carroll et al [26] developed and tested the psychological intervention “Group Therapy for Improving Emotional Acceptance and Regulatory Skills in Adults with ADHD” (GEARS), which consists of 14 weekly group sessions. In an uncontrolled pilot study with 226 participants, both treatment credibility and treatment satisfaction were rated as high, and preliminary clinical effects showed a reduction in emotion dysregulation symptoms with large effect sizes [26]. These findings are also supported by other studies [23,25,27-29]. On the other hand, Halmøy et al [24] did not find any significant difference in measures of emotion regulation between the control group and the treatment group in a randomized controlled trial of a 14-week dialectic behavioral therapy–based intervention. With the current evidence, it is thus premature to conclude whether psychological interventions are effective for adults with ADHD with co-occurring emotion dysregulation.

Digital psychological interventions have become increasingly popular in the past decade. The inclusion of digital tools in treatment interventions for adults with ADHD may be useful in addressing common challenges such as forgetfulness, nonadherence to treatment, and incomplete homework assignments [31,32]. For example, content from face-to-face therapy sessions, such as coping skills or psychoeducation, could be made available on the web through a website or a mobile app. The increased accessibility of such treatment elements may facilitate the generalization of therapeutic skills in everyday life for the clients, which is a central aim of most psychological interventions [33]. Blended interventions, where elements from face-to-face and digital treatment formats are combined, could be advantageous as they use the strengths of both treatment formats [33]. A systematic review of blended...
interventions in mental health care suggests that these interventions may save clinician’s time, lead to lower dropout rates, and help to maintain positive changes made in psychotherapy over time [34]. However, there have been few studies examining a blended treatment format among adults with ADHD. To our knowledge, there has been 1 randomized controlled trial examining the effect of a mobile app to deliver psychoeducation in a group-based intervention for adults with ADHD [35]. In this study, it was found that the participants who used the mobile companion app had a greater reduction in ADHD symptoms and a higher completion rate of homework assignments as compared to participants who received a printed version of the psychoeducation [35]. As such, this study provides evidence that psychological interventions for adults with ADHD may be augmented by the implementation of digital tools.

The overall aim of this study was to describe the development of and assess the feasibility and preliminary clinical effects of a blended psychological intervention for adults with ADHD. This intervention integrates face-to-face group sessions with a digital companion app for skills training, designed to address emotion regulation. The developmental process and the core content of the intervention are outlined in the methods section. Through the feasibility study of the intervention, the following research questions were addressed:

1. How do participants with ADHD rate their satisfaction and the credibility of the intervention?
2. What is the participants’ adherence level to the intervention?
3. What are the preliminary clinical effects of the intervention on emotion regulation, inattention, hyperactivity-impulsivity, quality of life, anxiety, depression, and executive functioning?

Methods

Phase 1: Development of Intervention

Participants
The participants who took part in the development phase of the intervention included 5 adults with ADHD, 2 clinical psychologists, 1 clinical psychology student, 2 human-computer interaction experts, and 2 user experience designers. The adults with ADHD were originally recruited from the local ADHD patient association. The adults with ADHD received a gift certificate worth NOK 400 (US $37) in total for their participation in the design workshop and evaluation meeting.

Methods of Development
The intervention was developed over 2 years through an iterative process. Multiple methods were used during this development (Table 1 shows an overview of methods and findings).
Table 1. Overview of the methods applied in the development of the intervention and key findings in the 2-year developmental process.

<table>
<thead>
<tr>
<th>Method</th>
<th>Content</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synthesis of quantitative</td>
<td>A synthesis of previous studies examining psychological interventions targeting emotion dysregulation in ADHD (^a) was conducted. In this step, we examined the format, treatment approach, and common psychotherapeutic elements in the interventions.</td>
<td>• Various psychological frameworks have been used in interventions targeting emotion dysregulation in ADHD, including DBT (^b), cognitive behavioral therapy, goal management training, and mentalization-based therapy.</td>
</tr>
<tr>
<td>literature</td>
<td></td>
<td>• A group-based format has been applied in previous interventions.</td>
</tr>
<tr>
<td>Synthesis of qualitative</td>
<td>A synthesis of previous qualitative examining the experience of participating in psychological intervention for adults with ADHD was conducted. In this step, we sought to understand the needs and preferences of psychological interventions among adults with ADHD.</td>
<td>• Most previous interventions apply a structured and manualized format.</td>
</tr>
<tr>
<td>literature</td>
<td></td>
<td>• Most previous interventions use homework assignments to generalize skills.</td>
</tr>
<tr>
<td>Expert meetings</td>
<td>The development process included several expert meetings. In these meetings, the format and elements of the intervention were discussed and reviewed. Suggestions regarding intervention content and design were made based on clinical expertise and experience.</td>
<td>• Common intervention elements were psychoeducation, mindfulness, acceptance strategies, self-monitoring, emotion regulation skills, behavioral analysis, planning and organization strategies, communication skills, and problem-solving skills.</td>
</tr>
<tr>
<td>Co-design workshop</td>
<td>The workshop included 5 adults with ADHD, 2 clinical psychologists, and 2 HCI (^c) experts. In the workshop, the adults were given information about the project and a general idea of the intervention. Following this, the adults were asked about challenging situations in terms of emotion dysregulation, common coping strategies, and their preferences for intervention content and features.</td>
<td>• In previous interventions, the opportunity to share experiences with peers was perceived as valuable by adults with ADHD.</td>
</tr>
<tr>
<td>Design sprint of companion app</td>
<td>A design sprint lasting 5 days was conducted to create a prototype of the companion app. An HCI expert led the design sprint, with the inclusion of 2 UX (^d) designers and 1 clinical psychologist as participants.</td>
<td>• Adults with ADHD report that they prefer an emphasis on strengths and solutions in treatment.</td>
</tr>
<tr>
<td>Cocreation of intervention</td>
<td>The first author created the intervention manual based on the previous findings. The protocol was then reviewed and revised by 2 other clinical experts, including 1 expert with ADHD.</td>
<td>• Incorporation of digital tools in treatment was perceived as useful by adults with ADHD.</td>
</tr>
<tr>
<td>manual</td>
<td></td>
<td>• Adult meetings with adults with ADHD.</td>
</tr>
<tr>
<td>Evaluation seminar</td>
<td>The evaluation seminar included the same participants as the co-design workshop. In the first part of the seminar, a clinical psychologist presented the content of 8 group sessions and the participants could give feedback. In the second part, an HCI expert presented a walk-through of the companion app, and participants could give feedback.</td>
<td>• Emotion dysregulation was a common challenge among adults with ADHD.</td>
</tr>
<tr>
<td>with adults with ADHD</td>
<td></td>
<td>• Perceived useful coping strategies by adults with ADHD: acceptance, self-compassion, distraction, reappraisal, time-out, and relaxation.</td>
</tr>
<tr>
<td>Consultations with software</td>
<td>We conducted several consultations with the software company that had the technical infrastructure for the companion app.</td>
<td>• Useful features in companion app: reminders, note-taking, overview of coping strategies, calendar, peer support, and inclusion of videos.</td>
</tr>
<tr>
<td>company</td>
<td></td>
<td>• Various intervention approaches, including DBT, cognitive behavioral therapy, goal management training, and mentalization-based therapy.</td>
</tr>
</tbody>
</table>

\(^a\)ADHD: attention-deficit/hyperactivity disorder.  
\(^b\)DBT: dialectic behavioral therapy.  
\(^c\)HCI: human-computer interaction.  
\(^d\)UX: user experience.
The Emotion Regulation Intervention for ADHD

Overview

The development process led to the “Emotion Regulation Intervention for ADHD” (ERIA), which is a structured psychological intervention aimed at improving emotion regulation skills in adults with ADHD. The intervention is manual-based and includes components from dialectic behavioral therapy skills training and positive psychology. ERIA consists of 8 face-to-face group sessions and a digital companion app for skills training in between sessions. Table 2 contains an overview of the intervention content in ERIA.

Table 2. Overview of the intervention, including session themes, content for group sessions, weekly homework, and skills training assignments.

<table>
<thead>
<tr>
<th>Session</th>
<th>Content</th>
<th>Homework and skills training</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td>• Introduction to group members and therapists</td>
<td>• Goal setting</td>
</tr>
<tr>
<td></td>
<td>• Overview of the program</td>
<td>• Identifying own strengths</td>
</tr>
<tr>
<td></td>
<td>• ADHD(^a) and emotion</td>
<td>• Complete plan for homework</td>
</tr>
<tr>
<td></td>
<td>• Discussion on challenges and strengths of ADHD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Strategies for homework completion</td>
<td></td>
</tr>
<tr>
<td>2. Mindfulness I</td>
<td>• Mindfulness practice</td>
<td>• Practice “what” skills in mindfulness (observation, description, and participation)</td>
</tr>
<tr>
<td></td>
<td>• Homework discussion</td>
<td>• Skills training log</td>
</tr>
<tr>
<td></td>
<td>• Mindfulness</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Introduction to “what skills” in mindfulness</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Exercise: “observe object in room”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Exercise: “describe the pictures”</td>
<td></td>
</tr>
<tr>
<td>3. Mindfulness II</td>
<td>• Mindfulness practice</td>
<td>• Practice “how” skills in mindfulness (being nonjudgmental, doing 1 thing at the time, and doing what works)</td>
</tr>
<tr>
<td></td>
<td>• Homework discussion</td>
<td>• Skills training log</td>
</tr>
<tr>
<td></td>
<td>• Common barriers for homework completion</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Introduction to “how skills” in mindfulness</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Exercise: “judgmental vs nonjudgmental claims”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Exercise: “multitasking”</td>
<td></td>
</tr>
<tr>
<td>4. Emotion regulation I</td>
<td>• Mindfulness practice</td>
<td>• Practice emotion regulation skills (observation of emotions and naming and describing emotions)</td>
</tr>
<tr>
<td></td>
<td>• Homework discussion</td>
<td>• Emotion diary</td>
</tr>
<tr>
<td></td>
<td>• Emotions and emotion regulation</td>
<td>• Skills training log</td>
</tr>
<tr>
<td></td>
<td>• Emotion regulation skills: part I</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Introduction to the emotion diary</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Exercise: “linking situations to emotions”</td>
<td></td>
</tr>
<tr>
<td>5. Emotion regulation II</td>
<td>• Mindfulness practice</td>
<td>• Practice emotion regulation skills (check the facts, opposite action, and problem-solving)</td>
</tr>
<tr>
<td></td>
<td>• Homework discussion</td>
<td>• Emotion diary</td>
</tr>
<tr>
<td></td>
<td>• Adaptive vs maladaptive regulation strategies</td>
<td>• Skills training log</td>
</tr>
<tr>
<td></td>
<td>• Emotion regulation skills: part II</td>
<td></td>
</tr>
<tr>
<td>6. Emotion regulation III</td>
<td>• Mindfulness practice</td>
<td>• Practice emotion regulation skills (covering basics needs and planning for challenging situations)</td>
</tr>
<tr>
<td></td>
<td>• Homework discussion</td>
<td>• Planning positive activities</td>
</tr>
<tr>
<td></td>
<td>• Emotion regulation skills: part III</td>
<td>• Emotion diary</td>
</tr>
<tr>
<td></td>
<td>• Tips for planning and organization</td>
<td>• Skills training log</td>
</tr>
<tr>
<td>7. Crisis management</td>
<td>• Mindfulness practice</td>
<td>• Practice skills in crisis management (stop and check-in, physical exercise, cold water, muscle relaxation, breath work, and distraction).</td>
</tr>
<tr>
<td></td>
<td>• Homework discussion</td>
<td>• Skills training log</td>
</tr>
<tr>
<td></td>
<td>• Skills for crisis management and intense emotions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Discussion on distraction strategies</td>
<td></td>
</tr>
<tr>
<td>8. Summary</td>
<td>• Mindfulness practice</td>
<td>• N/A(^b)</td>
</tr>
<tr>
<td></td>
<td>• Homework discussion</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Self-compassion</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Exercise: “supportive words”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Summary of program</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Road ahead: maintaining change and setbacks</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)ADHD: attention-deficit/hyperactivity disorder.  
\(^b\)N/A: not applicable.
**Group Sessions**

ERIA comprises 8 weekly group sessions. The group sessions are closed, and each group includes 6-8 adults with ADHD. In this study, the groups were led by 2 clinical psychologists with a minimum of 3 years of clinical experience. Each group session lasts approximately 1.5 hours, divided into 2 segments with a 15-minute break interlude. All sessions begin with a brief mindfulness exercise, led by 1 of the 2 psychologists. Afterward, group members are encouraged to share their reflections on the exercise. Following this, the group members share their experiences with the previous week’s homework assignment. After the break, the lead psychologist presents relevant theoretical and psychoeducational information and the new skills for the participants to practice at home. There are also discussion breaks and some practical exercises incorporated in the presentation of new skills.

**Companion App**

The participants are given access to a companion app, which they are asked to use for skills training at home between the group sessions. This app is organized with modules, where new skills training modules are released on a weekly basis following the course of the group sessions. The skills training modules include the weekly skills that the participants are to practice at home. In addition, the participants can use the companion app to message the group leader, log their skills training sessions, and access the PowerPoint presentations for the group sessions, as well as other relevant resources (eg, relevant websites or scientific papers on the weekly theme). The participants also receive an SMS text message reminder when the skills training modules are available and another SMS text message reminder if they have not accessed the module within 2 days. Given that this was a feasibility study, the app did not encompass all features originally present in its prototype. For instance, a peer-support feature suggested during the co-design workshop was excluded due to its resource-intensive nature. If the feasibility study yields encouraging outcomes, we intend to incorporate more sophisticated features in the app.

The app was a web-based application that could be accessed both through a computer and a smartphone. To access the app, the participants had to use the Norwegian web-based authentication platform, BankID, for a secure 2-factor log-in. BankID is a widely used platform for electronic identification for services within banking, health care, and education in Norway. The companion app was hosted on the eHealth platform developed by the software company “Youwell,” which is partnered with the Western Norway Regional Health Authority for clinical use and research. The platform has a patient and therapist portal, where participants can access the app through the patient portal, and the therapist can set up the program and monitor their patient’s progress in the therapist portal. The platform is also used for building the app and allows for the input of content, such as text, audio, videos, and images, to create modules. It has previously been used for other internet-delivered programs, including programs for ADHD [36], social anxiety [37], and cognitive residual symptoms after depression [38].

Figure 1 shows screenshots of the companion app. The first screen from the left shows the main page with an overview of the modules. The first box from the left reads, “Week 4: Emotion regulation I. This week we will practice skills in emotion regulation.” The second screen shows the instructions to a mindfulness skill with the text “Choose an option below and take 5 minutes to describe this in a nonjudgmental way.” The circles show 6 options they can choose from, which include describing (1) own emotions, (2) bodily sensations, (3) a tree or plant, (4) people walking by, (5) an object in the room, or (6) own thoughts. The third screen shows the side menu that participants can use to navigate within the module. This screen shows the mindfulness module and the side menu includes the specific module pages: “Welcome back,” “Skills,” “Observation,” “Description,” “Participate,” “Skills diary card,” and “Resources.”
Phase 2: Feasibility Study

Participants

Eligible participants for this feasibility study were adults with ADHD living in Bergen, Norway. The inclusion criteria for the study were as follows: (1) minimum age of 18 years; (2) a diagnosis of ADHD according to the DSM-5 (Diagnostic and Statistical Manual of Mental Disorders [Fifth Edition]) criteria; (3) current problems with emotion dysregulation as indicated by a score of ≥80 on the Difficulties in Emotion Regulation Scale (DERS); (4) having a smartphone or computer to access the companion app; and (5) the ability to attend in-person group sessions in Bergen, Norway. Exclusion criteria were (1) a high risk of suicidality, as indicated by having attempted suicide within the past year, having previously attempted suicide and reporting current suicidal ideations, or reporting current suicidal ideations and having a plan and preferred method; (2) co-occurring severe mental illness, including substance abuse disorder, psychosis, and major depressive disorder; and (3) current participation in another psychological treatment intervention. However, participants could still partake in the study if they had less severe psychiatric conditions, such as mild to moderate anxiety or depression, and if they received pharmacological treatment for their ADHD or other conditions.

Recruitment

Participants were recruited through the ADHD patient advocacy group “ADHD Norge,” which shared information about the study with its members through email and social media. The study opened for participants to sign up on November 24, 2022, and closed within a week (November 30, 2022) due to a large number of individuals signing up for the study. Interested participants signed up through a website that contained a screening survey to examine eligibility as well as information about the study. Eligible participants were contacted for a face-to-face screening with a clinical psychologist. During this screening, eligibility in terms of psychiatric comorbidities and suicidality was examined using the Mini International Neuropsychiatric Interview [39]. Participants were also asked about the date, clinic, and diagnosing clinician for their ADHD diagnosis. In addition, the participants were asked open questions about ADHD symptoms (ie, could you tell me about your current ADHD symptoms?) and everyday functioning (ie, how do you experience that ADHD affects you in your daily life?). The participants had to report symptoms and functional impairments that were in accordance with the DSM-5 criteria for ADHD, as assessed by a clinical psychologist. Those who were deemed eligible for participation following this screening were invited to take part in the study and to sign an informed consent form through the companion app.

Outcome Measures

Overview

The participants were given the preassessment 1 week before the intervention started and the postassessment 1 week after the intervention ended. The assessments were completed on the internet, except for the Behavior Rating Inventory of Executive Functioning–Adult version (BRIEF-A), which was given in-person at the first and last group session.

Credibility

The third item of the Credibility and Expectancy Scale (CEQ) was used to examine treatment credibility of the intervention [40]. The item states, “Would you recommend this treatment to a friend with similar challenges?” and the responses are given on a scale from 1 (not certain at all) to 9 (very certain).
Treatment Satisfaction
The participants were asked about whether they thought they would continue using the coping skills they had learned in the future and responded on a 4-point scale from 1 (not very likely) to 4 (very likely). They were also asked about which elements of the intervention they considered to be most useful (multiple choice with option to add own text). The participants were also asked about negative experiences with the intervention (yes or no), which they could elaborate on in an open-text field.

Adherence
Adherence was assessed by the number of group sessions attended and the number of completed skills training modules in the companion app. Participants that attended at least 6 of 8 group sessions were defined as treatment completers. The participants were also asked about how many days per week they practiced the coping skills.

Difficulties in Emotion Regulation Scale
DERS is a self-report questionnaire that is commonly used to assess emotion dysregulation in clinical populations [41]. The scale includes 36 items rated on a 5-point Likert scale ranging from 1 (never) to 5 (almost always), yielding a total score between 36 and 180, with 180 indicating the most severe problems with emotion dysregulation.

The Adult ADHD Self-Rating Scale
The Adult ADHD Self-Rating Scale (ASRS) is a self-report questionnaire that is used to assess symptoms of inattention and hyperactivity-impulsivity [42]. The scale includes 18 items, with 9 items reflecting inattentive symptoms and 9 items reflecting hyperactive-impulsive symptoms. Responses are given on a 5-point scale with options 0 (never), 1 (rarely), 2 (sometimes), 3 (often), or 4 (very often), giving a total score between 0 and 72 and a score between 0 and 36 for the inattention and hyperactivity-impulsivity subscales.

The Adult ADHD Quality of Life Measure
The Adult ADHD Quality of Life (AAQoL) measure is used to assess quality of life among adults with ADHD [43]. The scale includes 29 items rated on a scale from 1 (not at all or never) to 5 (extremely or very often), yielding a total score between 0 and 100.

Hospital Anxiety and Depression Scale
The Hospital Anxiety and Depression Scale (HADS) is a self-report questionnaire used to assess symptoms of depression and anxiety [44]. The scale includes 14 items, with 7 items reflecting anxiety symptoms and 7 items reflecting depressive symptoms. The response options range from 0 to 3, with 3 being the most severe level. The scale yields a total score between 0 and 42 and a score between 0 and 21 for the anxiety and depression subscales.

The Behavior Rating Inventory of Executive Functioning
BRIEF-A is a self-report questionnaire used to assess executive functioning in everyday life [45]. The scale consists of 75 items, which are rated on a 3-point scale (1=never, 2=sometimes, and 3=often). For this study, we report the Global Executive Composite score, which is an overall summary score including 9 clinical BRIEF-A subscales.

Statistical Analysis
SPSS software (IBM Corp) was used for all statistical analyses [46]. The participant demographics, adherence measures, and treatment credibility measures were assessed using descriptive statistics, which include calculation of means, frequencies, ranges, and SDs. To evaluate preliminary clinical outcomes, paired sample t tests were used with an initial significance level set at .05. Due to the risk of family-wise error (type I error) associated with multiple t tests, a Bonferroni correction was included. This adjustment was achieved by dividing the α level by the number of conducted hypotheses tests, that is, t tests (.05/8), resulting in a corrected significance level of .006. The choice of analytic approach necessitated the inclusion of cases with both pre- and postassessment. The magnitude of treatment effect was quantified using standardized effect sizes, estimated through Cohen’s d, with the formula (M_2 – M_1)/SD_pooled. The pooled SD was calculated by \( \sqrt{SD_1^2 + SD_2^2} \). Effect sizes were interpreted according to conventions, where \( d=0.20 \), \( d=0.50 \), and \( d=0.80 \) are defined as small, moderate, and large, respectively.

Ethical Considerations
The study was reviewed and approved by the Regional Ethics Committee of Norway, Region West (494659). The participants were informed about the study and their rights both in-person and in writing. Before participation, all participants signed an informed consent form, acknowledging that they could withdraw their consent at any time without any repercussions. The data were pseudoanonymized and stored on a dedicated research server according to regulative standards at the university and hospital. The participants were compensated NOK 1000 (US $90) for their participation in the study.

Results

Participants
A total of 16 participants took part in this study (Table 3 shows participant characteristics). The majority were diagnosed with ADHD in adulthood, with the mean age for receiving the diagnosis being 30.4 (SD 11.7) years.

All participants were recruited within 1 week. During the recruitment period, 68 adults completed the digital prescreening, of whom 92.6% (n=63) met the inclusion criteria and provided contact information (Figure 2). However, due to the limited number of study places in this feasibility study, we only assessed the first 27 individuals who signed up for the phone screening, as this was sufficient to reach the maximum capacity of 16 included participants. We sent an SMS text message to the remainder of the people who signed up, informing them that the study was fully booked. To ensure a more balanced sex distribution in the study, 6 study places were reserved for men, as we wanted to make sure we received feedback from both sexes.
Table 3. Participant characteristics in the feasibility study (n=16).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>10 (62.5)</td>
</tr>
<tr>
<td>Male</td>
<td>6 (37.5)</td>
</tr>
<tr>
<td><strong>Age group (years)</strong></td>
<td></td>
</tr>
<tr>
<td>18-34</td>
<td>7 (43.8)</td>
</tr>
<tr>
<td>35-44</td>
<td>7 (43.8)</td>
</tr>
<tr>
<td>45-55</td>
<td>1 (6.2)</td>
</tr>
<tr>
<td>55-64</td>
<td>1 (6.2)</td>
</tr>
<tr>
<td><strong>Educational level</strong></td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>8 (50)</td>
</tr>
<tr>
<td>College or university level</td>
<td>8 (50)</td>
</tr>
<tr>
<td><strong>Occupational status</strong></td>
<td></td>
</tr>
<tr>
<td>Employed or student</td>
<td>9 (56.2)</td>
</tr>
<tr>
<td>Work assessment allowance or sick leave</td>
<td>6 (37.5)</td>
</tr>
<tr>
<td>Disability pension</td>
<td>1 (6.2)</td>
</tr>
<tr>
<td><strong>ADHD medication status</strong></td>
<td></td>
</tr>
<tr>
<td>Daily medication usage</td>
<td>13 (81.3)</td>
</tr>
<tr>
<td>Weekly or monthly usage</td>
<td>2 (12.4)</td>
</tr>
<tr>
<td>Rarely, a few times a year</td>
<td>1 (6.2)</td>
</tr>
</tbody>
</table>
In terms of adherence, a total of 3 participants dropped out of the study before starting the intervention, making a pretreatment dropout rate of 18.8% (3/16). Among the participants who received the intervention (n=13), the average number of group sessions attended was 6.2 out of 8 sessions. A total of 4 (26.6%) participants were treatment dropouts, defined as participants who attended less than 6 group sessions. Consequently, the cumulative dropout rate, including the participants who dropped out before starting the intervention, reached 43.7% (n=7). Regarding the use of the companion app, the participants had a mean completion of 6.8 out of 8 modules. On average, the participants reported practicing the skills 4.3 days per week.

The intervention was generally well received in terms of treatment credibility with a mean rating of 7.1 (SD 2.6; range 1-9). Among the 10 participants who completed the postassessment, 7 participants reported that they were very certain or certain that they would recommend the treatment to a friend facing similar challenges as themselves; 2 were somewhat certain, whereas 1 was not at all certain about recommending it.

Feedback from participants highlighted overall satisfaction with the intervention, with a mean rating of 3.3 (SD 0.9; range 1-4). All but 1 participant planned to continue using the learned skills. The participants rated meeting others with ADHD (n=10), the in-person group sessions (n=7), the skills (n=7), and therapist support (n=4) as the most useful elements of the intervention. Their feedback also included suggestions for improvement, with 3 participants recommending more time for group interactions and discussions among the group members. Regarding this, 1 participant suggested that the group members should be able to interact between the sessions through the companion app. Other suggestions for improvements included an extension of the program by adding more group sessions, making the companion app available for direct download, incorporating more reminders, providing participants with a printed version of the skills, and involving an individual with ADHD as a group presenter for skills demonstration and experience sharing. While feedback was largely positive, 2 participants expressed negative experiences with the program: 1 found the skills in the crisis management module to be inadequate in emotional crises, and another felt that the intervention was not sufficiently tailored to ADHD. Importantly, there were no reports of clinical deterioration among the participants.
Clinical Outcomes

There was an overall statistically significant decrease in self-reported emotion dysregulation from before to after treatment, with a strong effect size ($d=2.0$). Table 4 shows the individual scores for emotion dysregulation, indicating a change in the positive direction for all participants.

Overall, the group of participants showed a significant decrease in secondary clinical outcome scores, reflecting the level of inattention and hyperactivity-impulsivity from before to after assessment. While initial analyses showed a change in quality of life and depressive symptoms, these changes were nonsignificant with a Bonferroni correction. No significant improvement was found in measures of executive functioning or anxiety symptoms (Table 5 contains an overview of all preliminary clinical outcomes).

Table 4. Individual scores for the Difficulties in Emotion Regulation Scale from before to after assessment (n=10).

<table>
<thead>
<tr>
<th>Participant</th>
<th>Preassessment scores</th>
<th>Postassessment scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>130</td>
<td>101</td>
</tr>
<tr>
<td>2</td>
<td>138</td>
<td>84</td>
</tr>
<tr>
<td>3</td>
<td>110</td>
<td>93</td>
</tr>
<tr>
<td>4</td>
<td>147</td>
<td>91</td>
</tr>
<tr>
<td>5</td>
<td>96</td>
<td>83</td>
</tr>
<tr>
<td>6</td>
<td>130</td>
<td>94</td>
</tr>
<tr>
<td>7</td>
<td>107</td>
<td>96</td>
</tr>
<tr>
<td>8</td>
<td>139</td>
<td>117</td>
</tr>
<tr>
<td>9</td>
<td>133</td>
<td>95</td>
</tr>
<tr>
<td>10</td>
<td>107</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 5. Clinical outcomes of emotion dysregulation, inattention, hyperactivity-impulsivity, quality of life, depression, anxiety, and executive functioning from before to after assessment (n=10).

<table>
<thead>
<tr>
<th>Outcome measure</th>
<th>Preassessment, mean (SD)</th>
<th>Postassessment, mean (SD)</th>
<th>$P$ value(^a)</th>
<th>Effect size, Cohen $d$</th>
<th>Difference, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>DERS(^b)</td>
<td>123.7 (17.2)</td>
<td>95.4 (9.6)</td>
<td>&lt;.001</td>
<td>2.0</td>
<td>15.8 to 40.8</td>
</tr>
<tr>
<td>ASRS(^c) full scale</td>
<td>51.9 (9.6)</td>
<td>41.9 (8.9)</td>
<td>&lt;.001</td>
<td>1.1</td>
<td>6.2 to 13.8</td>
</tr>
<tr>
<td>ASRS inattention</td>
<td>26.9 (4.1)</td>
<td>22.3 (5.2)</td>
<td>&lt;.001</td>
<td>1.0</td>
<td>2.3 to 6.9</td>
</tr>
<tr>
<td>ASRS hyperactivity-impulsivity</td>
<td>25.0 (6.4)</td>
<td>19.6 (5.1)</td>
<td>.002</td>
<td>0.9</td>
<td>2.5 to 8.3</td>
</tr>
<tr>
<td>AAQoL(^d)</td>
<td>44.4 (9.6)</td>
<td>56.7 (11.8)</td>
<td>.03</td>
<td>e</td>
<td>–23.3 to –1.3</td>
</tr>
<tr>
<td>HADS(^f) Anxiety</td>
<td>10.8 (4.3)</td>
<td>9.8 (3.6)</td>
<td>.24</td>
<td>—</td>
<td>–0.8 to 2.8</td>
</tr>
<tr>
<td>HADS Depression</td>
<td>6.0 (3.7)</td>
<td>4.3 (3.2)</td>
<td>.03</td>
<td>—</td>
<td>0.2 to 3.2</td>
</tr>
<tr>
<td>BRIEF-A(^g) GEC(^h)</td>
<td>154.5 (15.9)</td>
<td>150.8 (12.6)</td>
<td>.20</td>
<td>—</td>
<td>–7.8 to 15.1</td>
</tr>
</tbody>
</table>

\(^a\)Significance level set to .006 with Bonferroni correction.
\(^b\)DERS: Difficulties in Emotion Regulation Scale.
\(^c\)ASRS: Adult ADHD Self-Rating Scale.
\(^d\)AAQoL: Adult ADHD Quality of Life.
\(^e\)Not available.
\(^f\)HADS: Hospital Anxiety and Depression Scale.
\(^g\)BRIEF-A: Behavior Rating Inventory of Executive Functioning.
\(^h\)GEC: Global Executive Composite.

Discussion

Principal Findings

The aim of this study was to describe the development and assess the feasibility of ERIA, a blended digital and face-to-face intervention targeting emotion dysregulation in adults with ADHD. Overall, the findings were promising and supported the feasibility of ERIA. Both treatment satisfaction and credibility were generally good, which aligned with findings from other psychological interventions targeting emotion dysregulation in adults with ADHD [26]. More specifically, the participants emphasized the group component and meeting others with...
ADHD as useful aspects of the intervention. This is in line with previous studies showing that providing a forum where one can share experiences with peers is particularly valuable in treatment settings for this group of adults [47]. While the results are promising, there is room for refinements, including an even better tailoring of the intervention to ADHD. An important step forward would be to involve adults with ADHD in the review and refinement process to ensure that all key aspects of the interventions are well-adapted to the challenges and difficulties facing adults with ADHD. Nonetheless, the heterogeneous nature of ADHD requires finding a balance between including intervention components that fit most participants and including specialized strategies addressing the specific needs of a subset of individuals with ADHD.

This study highlighted concerns regarding treatment adherence, especially regarding pretreatment dropout. A total of 3 participants dropped out before starting the intervention, while another 4 participants did not complete the intervention. This finding aligns with well-known challenges related to treatment adherence among adults with ADHD [48,49]. The issues related to adherence in this study may be partially attributed to our community-based sample. Motivation for completing psychological treatment may thus have been lower than among clinic-recruited participants. However, there could also be intervention-specific factors related to dropouts, such as dissatisfaction with the intervention or finding the intervention too demanding. Yet it is also worth noting that although high levels of treatment adherence are generally seen as favorable for the patient, treatment dropout does not necessarily equate to clinical failure for the patients themselves. Still, when reviewing the current intervention, it may be necessary to include more strategies to prevent early termination of treatment. A literature review on the topic found that strategies such as pretherapy preparation, patient selection, time-limited treatment contracts, appointment reminders, motivation enhancement, facilitation of a therapeutic alliance, and facilitation of affect expression were specific strategies that could be applied to reduce premature termination of treatment across different psychiatric disorders [50]. As we progress, understanding the multifaceted factors associated with dropout will be crucial.

With regard to the companion app, it is interesting to note that the participants who initiated the intervention generally completed a high number of skills training modules in the companion app, with the majority completing all modules. This finding is in line with the results from Selaskowski et al [35], which showed that the inclusion of a mobile app in a group-based intervention for adults with ADHD was linked to higher homework compliance [35]. Given the importance of homework for behavioral change as well as the common challenges related to homework compliance in psychological interventions, facilitating methods to ensure high attrition of homework may be particularly useful [51]. However, it is important to note that with the design of this study, we cannot determine whether the companion app resulted in higher homework compliance as opposed to not including the app; this would be an interesting topic for future studies.

The results were promising in terms of preliminary clinical findings, with participants showing a significant and large reduction from pre- to post treatment in emotion dysregulation. These findings are in accordance with findings from other psychological interventions targeting emotion dysregulation in individuals with ADHD [26]. Significant improvements were also observed in the secondary clinical outcome measures, including inattention and hyperactivity-impulsivity. However, given the study’s small scale, nonrandomized, and uncontrolled design, the clinical findings must be considered with caution. Moreover, we found no significant changes in the measures of anxiety, depression, quality of life, or executive functioning. Regarding executive functioning, previous research has shown that measures of executive control remain stable in ADHD, regardless of remission or persistence [52]. More generally, the effect of interventions that are designed to improve cognitive abilities or executive functioning appears to be domain-specific and show mixed results [53,54].

Taken together, the results from this study were encouraging and call for further development of the intervention and a more extensive examination of clinical effects in a randomized controlled trial.

Limitations
This feasibility study has some limitations that should be noted, in particular the small sample size, the absence of a control group, and the lack of randomization. Due to these limitations, conclusions regarding the intervention’s clinical effects remain elusive, and we cannot rule out placebo or other random effects. Yet, the aim of this study was not to examine the effectiveness of the intervention but rather to assess its feasibility before paving the way for larger trials.

A further limitation is that 3 participants who took part in the intervention did not complete the postassessment. Therefore, it is possible that this could have impacted the treatment satisfaction and credibility scores of the interventions. Furthermore, the participants were recruited from the community and may therefore differ in some way from typical clinic-recruited adults with ADHD. The reliance on self-report scales to evaluate clinical outcomes in this study should also be considered a limitation. In future studies, it would be interesting to include other measures, such as those generated from sensor technology, clinician ratings, and performance on cognitive tests.

Conclusion
In conclusion, the results from this feasibility study support the potential of ERIA as a feasible intervention for addressing emotion dysregulation in adults with ADHD and call for further investigation in a randomized controlled trial. The blended approach, integrating digital and face-to-face elements, may offer some advantages compared to an exclusively digital or face-to-face treatment format. The in-person group sessions were especially valued because they provided opportunities to interact with peers. Meanwhile, the high completion rate of the companion app modules indicates their potential to facilitate skills training.
Acknowledgments

The study was funded by Helse Vest research funds. The authors would like to thank all the participants in this study and the experts to experience. We would also like to thank the patient advocacy group “ADHD Norge” for sharing information about this study. We would further like to thank Hedda Bakken, Sara Furuholmen, and Nanna Skram for their contributions to the project.

Authors’ Contributions

ESN, AJL, and FG contributed to the development, idea, and design of the study. ESN, AJL, SLA, and VS contributed to the data collection for the feasibility study. ESN conducted the statistical analyses, while AJL, FG, JK, SLA, and VS contributed to the interpretation of the results. ESN was responsible for the drafting of the manuscript, while AJL, FG, JK, SLA, and VS critically reviewed and revised the draft. All authors approved the final version of the manuscript and take full accountability for the work.

Conflicts of Interest

JK has delivered talks at educational events sponsored by Medice; all funds were received by King’s College London and used for studies of ADHD. AJL has received speaker fees and travel expenses from Medice.

References


Abbreviations

AAQoL: Adult ADHD Quality of Life
ADHD: attention-deficit/hyperactivity disorder
ASRS: Adult ADHD Self-Rating Scale
BRIEF-A: Behavior Rating Inventory of Executive Functioning–Adult version
CEQ: Credibility and Expectancy Scale
DERS: Difficulties in Emotion Regulation Scale
DSM-5: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
ERIA: Emotion Regulation Intervention for ADHD
GEARS: Group Therapy for Improving Emotional Acceptance and Regulatory Skills in Adults with ADHD
HADS: Hospital Anxiety and Depression Scale

Edited by A Mavragani; submitted 25.10.23; peer-reviewed by M Guth; comments to author 12.12.23; revised version received 17.12.23; accepted 17.12.23; published 17.01.24.

Please cite as:
A Blended Intervention Targeting Emotion Dysregulation in Adults With Attention-Deficit/Hyperactivity Disorder: Development and Feasibility Study
JMIR Form Res 2024;8:e53931
URL: https://formative.jmir.org/2024/1/e53931
doi:10.2196/53931
PMID:38231536

©Emilie S Nordby, Frode Guribye, Viktor Schønning, Sander Lindholm Andersen, Jonna Kuntsi, Astri J Lundervold. Originally published in JMIR Formative Research (https://formative.jmir.org), 17.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Designing a Web-Based Navigation Tool to Support Access to Youth Mental Health Services: Qualitative Study

Alison L Calear, BAppPsych (Hons), PhD; Philip J Batterham, BSc Hons, MPH, PhD; Sonia M McCallum, GradCertPHC, BHlthSC (Hons), BSc, PhD; Michelle Banfield, BA (Hons), BSc, PhD; Elizabeth Moore, GradCertHPE, BS, MB; Natalie Johnson, GradCertPM; Alyssa R Morse, AMusA, BPsych (Hons), PhD

1Centre for Mental Health Research, The Australian National University, Canberra, Australia
2ALIVE National Centre for Mental Health Research Translation, Melbourne, Australia
3Office for Mental Health and Wellbeing, ACT Health, Canberra, Australia

Corresponding Author:
Alison L Calear, BAppPsych (Hons), PhD
Centre for Mental Health Research
The Australian National University
63 Eggleston Road
Canberra, 2601
Australia
Phone: 61 261258406
Fax: 61 261250733
Email: Alison.Calear@anu.edu.au

Abstract

Background: Many young people with mental health problems do not readily seek help or receive treatment and support. One way to address low help-seeking behavior is to improve access to information on mental health services and how to navigate the mental health system via a web-based tool. Seeking input from the end users (young people and parents or caregivers) on key features of the tool is imperative to ensure that it is relevant, engaging, and likely to meet their needs and expectations.

Objective: This study aims to investigate young person and parent or caregiver views on the design, content, functioning, and user experience of a web-based mental health navigation tool to support connection to mental health services for children and young people aged up to 25 years.

Methods: A total of 4 online focus groups were conducted: 2 with young people aged 16 years and older (total n=15) and 2 with parents or caregivers (total n=13). Focus groups were structured around a series of guiding questions to explore participants’ views on content, features, user experience, and design of a mental health navigation website. Focus groups were audio recorded with detailed notes taken. In addition, 53 young people aged 16-25 years and 97 parents or caregivers completed an online survey, comprising closed- and open-ended questions; open-ended responses were included with the focus group data in the qualitative analysis. All qualitative data were analyzed using thematic analysis.

Results: A total of 2 topic areas and 7 themes were developed. The first topic area covered the types of information needs of young people and parents. Identified themes concerned the scope of the navigation website, as well as the provision of up-to-date and practical information on how to navigate the whole help-seeking process. The second topic area covered website features that would be beneficial and included the consideration of the website design; search engines; supported navigation; and forums, reviews, and user accounts.

Conclusions: This study provides important insights into the navigation needs of young people and parents or caregivers in seeking mental health services. Key findings identified through this research have directly informed the development of MindMap, a web-based youth navigation tool providing a searchable database of local services, including a clear description, their location, and potential wait times. The website can be navigated independently or with support.

(JMIR Form Res 2024;8:e48945) doi:10.2196/48945

KEYWORDS
mental health services; youth; navigation tool; mental health; website; user experience; design; service; services; access; accessibility; health care system


Introduction

Globally, mental disorders are common and often emerge during childhood and adolescence [1,2]. The worldwide pooled prevalence of mental disorders in children and adolescents has been reported to be between 12.7% and 13.4% [2,3], while suicide is one of the leading causes of mortality in this age group [4]. Timely access to appropriate services and supports for mental disorders and suicidal distress can mitigate the persistence of poorer health, academic, and social outcomes into adulthood [5].

Research suggests, however, that many young people do not readily seek or receive treatment or support for psychological distress, suicidal ideation, or suicidal behavior [3,6,7]. A recent systematic review and meta-analysis of the prevalence of mental disorders in children and adolescents in high-income countries found that only 44.2% of young people with mental disorders received any services for their conditions [3]. Another study conducted in Australia reported slightly higher rates of service use among children and adolescents, with 56% accessing services for emotional and behavioral problems [7]. However, Sawyer et al [8] reported that only 12% of Australian young people aged 6-17 years with a mental health condition received what was considered adequate treatment.

Several possible drivers of low help-seeking behavior among young people, parents, and other caregivers have been identified. A recent review found that limited mental health knowledge, embarrassment and perceived social stigma, a lack of perceived confidentiality and trust in mental health providers, financial costs, logistical barriers, and limited availability of services were common barriers to accessing professional help for mental health problems among young people [5]. Similar barriers were identified among parents seeking treatment for mental health problems in their children and adolescents, with a lack of knowledge of where or how to seek help and a limited understanding of the mental health system identified as key barriers to service use [9].

One way to address some of these barriers may be to provide young people and parents or caregivers with a navigation tool they could use to identify available services and traverse often complex and disconnected mental health systems. Similar to other high-income countries, there are multiple entry points into the mental health care system in Australia, but access is often dependent on the individual knowing about them. As such, being able to access all service options in the same place would overcome service knowledge barriers and save users’ time trying to identify what is available in their area. Providing such tools online through an interactive website would also make this information more accessible, as adolescents and parents report seeking mental health information online [10-13] and a website would allow greater reach and timely updating of service information.

In a survey assessing parents' help-seeking for their adolescent’s mental health, over 75% of participants indicated that they would use the internet to find information about services [14]. In developing such a website, it is imperative that young people and parents or caregivers are involved, as there is clear evidence in the literature that the involvement of end users is essential to ensuring uptake and that the website meets the needs and expectations of users [15,16].

This paper reports the outcomes of focus groups and online surveys conducted with young people, parents, and other caregivers to identify the design, content, functioning, and user experience of a web-based mental health services navigation tool for children and young people aged up to 25 years in the Australian Capital Territory (ACT). This study was conducted in partnership with the ACT Office for Mental Health and Wellbeing and provides unique insights into the navigation needs of this population, with a dearth of previous research in this area, and the importance of the consultation process in producing an interactive website that is relevant, engaging, and likely to meet the needs of end users.

Methods

Participants

A total of 4 online focus groups (N=28), with 5-10 participants each, were conducted. Of these focus groups, 2 focus groups (15/28, 54%; 10/15, 66% female) were conducted with young people aged 16 years and older. These groups were conducted with members of 2 existing youth mental health reference groups: the ACT Youth Advisory Council and the ACT headspace Youth Advisory Group, who had experience representing the wider interests and views of their peers. The other 2 focus groups involved parents or caregivers (13/28, 46%; 12/13, 92% female). Participants for these were drawn from expressions of interest from the community and included parents with and without mental health service experience. To attract diverse representation in the focus groups, we advertised for recruitment through ACT mental health sector newsletters and on the social media website Facebook. Facebook advertising targeted adults in the ACT who identified as a parent or caregiver of a child aged 10-25 years. The online modality for the focus groups was chosen to maximize ease of attendance and minimize potential COVID-19 exposure.

Alongside the focus groups, 150 people participated in an online survey to further garner the perspectives of young people and parents or caregivers. Participants were young people aged 16-25 years (53/150, 35%; 44/53, 83% female; mean age 20, SD 2.8 y) and the parents or caregivers (97/150, 65%; 90/97, 93% female; mean age 45.5, SD 7.6 y) of young people aged 0-25 years (0-6 y: 20/97, 21%; 7-11 y: 38/97, 39%; 12-15 y: 47/97, 48%; 16-18 y: 30/97, 31%; and 19-25 y: 37/97, 38%). Participants for the survey were recruited through Facebook advertising. The Facebook advertisements were delivered to 21,600 accounts, with 1300 people engaging with the Facebook post and 374 clicking the link and commencing the survey. Of those who commenced the survey, 40.1% (150/374) went on to complete and submit the survey for analysis. The majority of participants reported English as their first language (47/53, 89% young people and 93/97, 96% parents or caregivers) and had accessed pediatric mental health services in the past (45/53, 85% young people and 86/97, 89% parents or caregivers). Just over half (28/53, 53%) of the young people who completed the survey identified as lesbian, gay, bisexual, transgender, intersex, or other sexual orientation.

https://formative.jmir.org/2024/1/e48945

XSL•FO
RenderX
queer (LGBTIQ+), whereas 21% (11/53) reported having a disability.

**Ethical Considerations**

Ethical approval for this study was obtained from the Australian National University Human Research Ethics Committee (protocol 2020/200). All participants were provided with an information sheet outlining the study. Participants in the focus groups were required to return a signed consent form via email before attending the online focus group, whereas survey participants consented to participate in the survey by checking a box at the beginning of the survey. All survey data were collected anonymously, whereas all focus group data were deidentified before analysis and data storage. All focus group participants were provided with a small honorarium of AU $60 (US $40) in recognition of their time and contribution to the study.

**Materials**

The focus groups were structured, with a series of guiding questions and prompts that explored the views of young people and parents or caregivers on the potential design, content, functioning, and user experience (eg, optional phone and email support) of a web-based navigation tool (the full list of questions and prompts is given in Multimedia Appendix 1). Questions and prompts were developed in collaboration with staff members from the ACT Office for Mental Health and Wellbeing to gather information required to progress website development from an early concept and include youth and parent perspectives in the design process. Participants were encouraged to share their perspectives verbally with the group or to note any thoughts or ideas in the chat function.

The online survey assessed a range of topics relevant to the development of a youth navigation tool including website features (eg, information, links, quizzes, and service contact details), features and topics by which services could be searched (eg, cost, location, age, and gender), navigation support (eg, phone and online), and possible account functions (eg, acceptable account details and storage of information). Questions were developed with iterative feedback from the ACT Office for Mental Health and Wellbeing and the Youth Coalition of the ACT (peak body for youth affairs in the ACT) to ensure that they covered key issues for website design and were appropriate and accessible for young people. Participants were asked to share their preferences through a combination of close-ended and free-text responses.

**Procedure**

Because of the COVID-19 pandemic, all focus groups were conducted online in June 2020 using the Zoom videoconferencing platform (Zoom Video Communications). Given that this study was conducted in partnership with the ACT Office for Mental Health and Wellbeing, all focus groups were cofacilitated by a member of the research team and the ACT Office for Mental Health and Wellbeing. A second member of the research team was also present at all focus group sessions. Their role was to take notes, assist with linking participants with clinical support (if required), and monitor any discussion via the chat function, which was encouraged. All focus groups ran for approximately 90 minutes, online discussions were audio recorded, and any chat conversations were also saved for inclusion in the analysis. One author (ARM) listened to each focus group recording and took detailed point-by-point notes on the content of participant discussions. These were combined with researcher notes taken during the session and participants’ chat conversations for analysis. Because of this approach, all illustrative quotes were drawn from survey data. Participants were provided with help-seeking resources at the conclusion of each focus group, and clinical support or debriefing was available on request.

Two separate surveys were developed for young people and parents or caregivers and were administered online in June 2020 using Qualtrics online survey software (Qualtrics). Surveys took approximately 20 minutes to complete. All participants were provided with a list of help-seeking resources at the conclusion of the survey.

**Data Analysis**

Qualitative data were managed using the NVivo 12 software (QSR International). A thematic analysis [17,18] was conducted by 1 author (ARM) on focus group data notes, chat conversations, and free-text survey responses to identify and summarize the key topics and preferences raised by parents and young people within each area of questioning, while preserving the breadth and diversity of perspectives presented. ARM is a lived experience academic with personal experience of mental health service use and professional expertise in youth mental health and service evaluation. Analyses and developing categories were regularly discussed with other members of the research team, including researchers and ACT Office for Mental Health and Wellbeing staff members present at the focus groups, to test assumptions and clarify developing categories. A combination of deductive and inductive approaches to coding was applied. Before commencing data analysis, an initial broad coding framework was developed based on the key areas of questioning during the focus groups: content, features, user experience, and design. Focus group data were deductively coded within each category. As coding progressed, subcategories of describing specific areas of preference and new categories not represented by the deductive framework were developed inductively. A detailed coding framework was developed from the focus group data, including the original broad areas of questioning and inductively developed categories and subcategories. This framework was applied to free-text survey responses, adjusting the framework as necessary to adapt to new information. Finally, data were examined across categories to construct common themes in participants’ preferences. The key topic areas were summarized and grouped under 2 broad categories: the most important types of information for a service navigation tool and important website features.

**Results**

**Overview**

The overarching topic areas arising from the focus groups and surveys covered the types of information needs identified by young people and parents and the website features that would be beneficial to support the mental health of young people.

https://formative.jmir.org/2024/1/e48945

 JMIR Form Res 2024 | vol. 8 | e48945 | p.298

(page number not for citation purposes)
Themes collectively developed from the focus groups and surveys are described within these 2 areas. Data from the focus groups and surveys are integrated throughout this section. Where findings relate only to a specific group (parents or young people) or data collection format (focus group or survey), this is noted in the text.

**Types of Information**

**Up-to-Date Information About Services**

Preferences and needs for mental health service information were key topics of discussion in the focus groups and one of the most frequent topics raised in open-ended survey responses. Participants emphasized that it was important for the information on a mental health service navigation website to be regularly updated, with an indication of when updates last occurred, and for no information about listed services to be missing. Content should be relevant to the local region targeted by the website (ACT and surrounding region), as national-level websites can be difficult to navigate and regional content may better cover the full spectrum of services available in that area.

> Anything that is Canberra specific is great - wait times, contacts, prices, explaining which professionals do what. For general advice there are lots of national mh [mental health] services already. [Young person, survey participant]

> It would need to be regularly updated to match service changes. It would need to be all evidence based. [Parent, survey participant]

Current and accurate information about service wait times was particularly desirable for all participant groups. Families commonly experienced a lag between identifying that a young person had a problem and being able to access support services. Providing wait time information on a service navigation website could change the way families choose to access services and direct them to services with better availability. Knowing about wait times at the beginning of the help-seeking process would also inform young people and families about likely time frames for receiving support and enable them to plan ahead.

> I think the idea of including information about wait list times is very important - as this is usually critical at the time that you are seeking information and support. [Parent, survey participant]

> I think most people are primarily looking for immediate reassurance and advice. If they can get an idea on how long it might take to access different services this also gives them a time-frame so they can plan. [Parent, survey participant]

Young people highlighted the value of a navigation website in providing detailed information about services that would help them make informed choices. Information about accessibility was particularly important for young people. Specifically, cost and parental consent were identified as key barriers to service access, which could be partly addressed by providing detailed cost information (eg, service fees, rebates, and what benefits a child or family may be entitled to), indicating whether parental consent was required to use a service, and noting any relevant legal considerations. Young people also wanted to know if they would be able to physically access a service before they arrived and recommended that the website include information about wheelchair access (eg, presence of ramps and elevators), accessible and gender-neutral toilets, the languages spoken by staff members, whether tailoring was available for people who are vision- or hearing-impaired, and the age range targeted by a service. The content of a service navigation website should also specifically address issues faced by different minority groups, including LGBTIQ+ communities, people with disabilities, and migrant communities.

> Along with a list of mental health services, also ways of accessing them, with or without parental involvement and potential cost. [Young person, survey participant]

> Information specifically geared towards minority groups (BIPOC, LGBTQIA+, people with disabilities etc.). Having information about different mental health issues is great but it would be a mistake not to factor in the relationship between mental health and other aspects of people’s identities. [Young person, survey participant]

Parents wanted a navigation website that provided clear information about the kinds of services available in their area, including government and private services and general practitioners who specialize in youth mental health. The ideal website would provide specific information about how to contact services, referral pathways, referral requirements, and whether new referrals were being accepted. Details about service specializations and appointment options (eg, availability of telehealth appointments) were also indicated as helpful across both participant groups.

> Information about accessing public v [versus] private mental health services and explaining the difference. [Parent, survey participant]

> Make sure info on telehealth is clear on if it’s: via the internet text-based, internet video-call, via the telephone, via text message. Don’t just say “this place offers telehealth services.” [Young person, survey participant]

**How to Seek Help, and What Happens Next?**

Parents emphasized that a service navigation website should include practical information about how to navigate the whole help-seeking process—from recognizing signs and providing support for their child through to good questions to ask a general practitioner when seeking a mental health referral and how to advocate for and manage the ongoing care of their child. Real stories of how other families navigated mental health issues, practical step-by-step instructions, and diagrams were suggested as useful tools to facilitate access to appropriate services. Young people also wanted clear step-by-step instructions describing how to access help, including how to access services without parental involvement. They noted that finding and visiting a new location can be scary. Providing extra information about the service location, including a map, pictures of the building and the front door, and pictures of staff members, could make the process easier. Information about different types of therapies...
Participants noted that help-seeking resources should foster hope, rather than disappointment, if a service access attempt did not work out. For example, a parent who completed the survey suggested that a service navigation website could illustrate what a good mental health service arrangement looks like and let people know it is okay to try different clinicians and services if the first referral was not a good fit. Parents also suggested that a navigation website should include easily accessible information on how to recognize, respond to, and seek urgent help during a crisis situation. One parent suggested including a function that could directly link people to a crisis service if needed (eg, a crisis telephone service like Lifeline); however, other participants felt that most people would already know about commonly advertised crisis hotlines.

Information on what a good mental health service arrangement should look like - depending on individual need - and it is ok to not just stick with one person/service forever or choose none. [Parent, survey participant]

I think you need information clearly on the homepage about crisis care...and other urgent items that you don't want to have to sift through a website for. [Parent, survey participant]

**Defining Website Scope**

In all 4 focus groups and the online surveys, participants raised questions about the ideal scope of a service navigation website. Participants tended to agree that the website’s scope should not be too broad. However, there was uncertainty about whether a navigation website should provide information about mental health in general or only provide information about services. Some participants also expressed a preference for a broader focus on health and well-being, rather than limiting information to mental health services. Young people suggested that broader content could include information about career and employment guidance, coping with current events, mental health at school, and quick references for self-management strategies that could be used while waiting for services (eg, mindfulness, distress tolerance, and coping with panic attacks). Parents were interested in content about mental illnesses (including mood, anxiety, and eating disorders), how to recognize them in specific age groups, and information about common comorbidities and related issues like aggression and self-harm.

This is far too mental illness and mental health service focused than I would be wanting. As a parent it’s helpful to have that information but I would also like resources that are tailor made for the site and have a focus on more on early intervention and mental wellbeing. [Parent, survey participant]

Providing links to different websites or existing online programs was suggested as an acceptable option to prevent the scope of a service navigation website from becoming too broad. For example, a navigation website could provide links to existing early intervention and mental well-being resources to support young people’s and parents’ well-being and assist parents to provide support when issues were first identified. Young people suggested providing links to research papers; stories about other people’s lived experience; and different forms of media that represent mental health in a productive way, such as video games, books, and movies. Participants noted that this approach could also connect a website to resources that fall beyond what they would typically define as a “mental health service,” including information about physical health, disability services, drug and alcohol services, and community programs and events that support and empower young people.

I would like to see it link not just mental health services, but other services to help a child overcome all the problems that may be adding to the mental health issue. For example, if the child is experiencing a lot of pain, if the child needs weight management help, if the child is being bullied or needs to develop resilience, if the child is on the autism spectrum, etc. [Parent, survey participant]

Stories from people who struggle with their mental well-being but have found support and renewed belief in themselves. [Young person, survey participant]

**Website Features**

**Website Design**

Participants generally agreed that website design would be an important element of a service navigation website’s success. Elements of website design highlighted by young people included a quality user interface that was easy to use and attractive to the intended audience. Young people described a well-designed website as colorful (but not too colorful or gimmicky), engaging, private, welcoming, local, and informed by what we know about young people and how they think. A level of seriousness in the design was required to ensure that the website was viewed as a reliable resource. Government-branded websites were viewed as a trustworthy source of local information but could be off-putting for young people who had previous negative experiences with government mental health services. A list of organizations that support the website could also signal the trustworthiness of information. Young people were also very aware of accessibility issues, recommending that a navigation website be designed to work across different platforms, for people with slow internet connections, and that it met relevant accessibility standards (eg, for people with low vision or lower literacy levels). A parent suggested that the name of the website was also important and needed to be inviting, explanatory, and nonstigmatizing.

Easy website navigation was another key design issue. When trying to communicate about one’s own mental health in a
difficult time, participants felt that the most important thing was to access information easily. Participants indicated that they would be more likely to visit a service navigation website for a specific need or to find specific information. The website content should be set out plainly, with clear pathways to the kind of information the user is looking for. Website design should not be overly complicated, avoiding the need to navigate through tabs and the presence of too many distractions (eg, moving images or videos). Young people were described as multitaskers; thus, a service navigation website needed to capture their attention quickly. To achieve this, it would be helpful to present information clearly and concisely, and to avoid walls of text that could be overwhelming. A frequently asked question section and fact sheets were seen as helpful, but only when presented as a dedicated webpage and not solely as a downloadable document.

### Search Engines

Parents and young people agreed that a service navigation website should systematically connect people to relevant services quickly. A good quality search engine and filtering system was a particularly important aspect of young people’s user experiences; the search bar can be the first port of call for young people trying to find help. A good search engine was described as easy to navigate, with tags and search terms updated as service information changes. In 1 focus group, young people suggested that a service navigation website could be designed like a nice online shopping experience, with tabs and subtabs for different categories of services and filters that allow users to refine their search and locate the most relevant services. However, 1 parent survey participant noted that a filter system would make them feel terrible if it indicated that there were no services matching their child’s needs, indicating that a balance between detail and generalizability may be required. Participants suggested that if there were no services meeting a young person’s search criteria, the website could direct them to resources, fact sheets, or other information they may find helpful.

...a search engine that finds services that are relevant to you. E.g. you could put in the tags “stressed” “self harming” “aged 17” “don’t want to involve parents” and the search engine would suggest; things you could do to help yourself, services you could access, and a helpline. [Young person, survey participant]

The filter would make me feel terrible if meant that my child had no services available so I think that you need to be careful regarding putting too many. [Parent, survey participant]

Focus group participants suggested having a quiz or questionnaire to help young people and parents navigate the website. For example, a pop-up box could appear when a person first accessed the website with some questions about what visitors are looking for. Answers could direct young people and parents to appropriate website sections, services, or self-help strategies. Parents suggested that a navigation website could include a symptom checklist, providing recommendations on whether a person needs to seek help from a health professional and within what time frame. Participants noted that any quizzes or checklists should be accompanied by a disclaimer stating that the website could not provide a diagnosis, and questions should be symptom or problem based, not diagnosis based. The results should be anonymous unless a young person chose to disclose them. However, some participants thought it would be important to determine if a young person was in immediate danger and requiring assistance. To facilitate help-seeking, a symptom checklist tool would ideally lead to an outcome, such as connecting users with a real person who can assist with identifying an appropriate service or next step.

Wouldn’t it be easier to complete a mental health survey upon entering the site that directs you to all the relevant pages? [Young person, survey participant]

Adolescence is a tricky time. Parents don’t know when to worry, when to escalate to professional help, and when to leave kids to muddle through. A quiz that helps navigate that would be super helpful. [Parent, survey participant]

### Supported Navigation

Participants were enthusiastic about having the option to contact a person, by phone or text-based chat, to help them gauge the seriousness of their issue, navigate the mental health system, connect with appropriate services, and answer questions about what to expect at an appointment. Parents emphasized the importance of creating a sense of trust, confidence, and reliability when a person makes a connection through a navigation website. Finding the right service could take a lot of time, research, and mistakes, particularly during times of stress. Some participants described help seeking as overwhelming, emphasizing the importance of positive experiences that could renew confidence in seeking help and support future service use. An interaction with a real person could be an opportunity to foster hope, positive regard, empathy, reassurance, and a sense of not being alone.

This is a great idea. I found navigating the system to be impossible to start with. I was googling everything, calling all these people, being passed from service to service and getting nowhere. None of the service providers knew what any of the other service providers could do. Someone to help you navigate that would be amazing - especially if you are out of your mind with worry and sleep deprivation like I was! [Parent, survey participant]

When asking questions about mental health problems, some young people reported preferring to talk to a health professional, whereas some parents and young people suggested that this would be a good opportunity for peers to support young people and carers and that young adult peers may be better able to connect with younger website users. However, 1 young person noted that talking to a healthy peer may be intimidating for some young people. Across the surveys and focus groups, participants suggested that the person they contacted should be genuine, engaged, supportive, friendly, empathetic, unhurried, unscripted, and well trained, with appropriate counseling skills to support stressed or distressed callers. Their understanding of
mental health and local health services should be broader than the understanding provided by the caller’s own experiences.

*Having an advisory line (telephone or chat) which is supportive, anonymous if wanted, friendly and unhurried, staffed by a real person who is quietly supportive but knowledgeable would be great to help people navigate the system and get a sense of where they are best placed to use their energies in pursuing or connecting with services.* [Parent, survey participant]

Participants emphasized that the purpose of any phone or text-based contact options and the roles of the people running them need to be very clear. Contact information should be clearly stated and easy to find, and ideally some contact options would be available outside of normal business hours (eg, evenings, nights, and weekends). Participants had concerns around potential privacy issues, particularly related to data collection and storage, and the need to support people who disclose thoughts of suicide and self-harm. Parents recommended that all interactions end with some kind of closure, for example, actions such as making an appointment for the person with an appropriate service or taking a concrete step that progresses the issue, with timely outcomes.

*I’d also be more comfortable in knowing how any conversations via the website were recorded and stored since there is always the chance of personal stuff coming up in conversations, and I wouldn’t want that to be accessible by anyone except those who are directly working to help me.* [Young person, survey participant]

Further, participants discussed the option of receiving a follow-up call or text, after interacting with a person via a service navigation website. Some participants felt that a follow-up call could be comforting for people who were currently on a waitlist, for example, by providing updates on wait times and identifying alternative sources of support. Survey participants were primarily interested in receiving follow-ups related to service access, for example, checking in to see if services had connected with a young person, how effective the service has been, and if the young person’s needs are being met. Although some participants also wanted the opportunity to give feedback about services, some parents felt oversurveyed by mental health services. Other participants felt that a follow-up would not be helpful in all circumstances. For example, 1 parent felt that receiving a phone call may just be more frustrating if accessing services was not going well. Participants agreed that any follow-up from a service navigation website should be opt-in, the user should have control over how and when they are contacted, young people should decide whether their parents are contacted, and any promises made by a navigation service should be honored (ie, call if you say you will call). Choice and control over contact was seen as particularly important for young people living in high-risk situations, where receiving a phone call or message could potentially be unsafe.

*I think having options is good, especially for kids in potentially dangerous situations. And I do think texting is popular with kids.* [Young person, survey participant]

*I think it is also important to see how well people went with actually accessing the services and whether their child’s needs were met and what negatives there were.* [Parent, survey participant]

**Forums, Reviews, and User Accounts**

Website functions that could allow users to share information were met with a mixed reception. The options discussed included forums, service reviews, and user accounts. Forums received the most positive reception, but with important safety considerations for implementation. They were described as a positive tool for young people and parents to connect with peers who have had similar experiences and share coping techniques or to connect with health professionals. However, to be safe and useful, a forum or chat room would require careful moderation. Participants suggested that forums could be provided by a service navigation website itself or the website could instead provide links to external, good quality, moderated mental health social media pages or similar services.

*Particularly when I was younger, online resources were huge! Whether that was headspace’s online counselling, or online chat forums with other people. I think those chats were absolutely fantastic, however they definitely needed expert moderation.* [Young person, survey participant]

*I’m a little concerned about forums and/or chat rooms. They would need to be carefully monitored to make sure there are no trolls responding negatively or people using cyber bullying.* [Parent, survey participant]

User reviews of mental health services were raised as a possibility, but this option had both advantages and disadvantages. Reviews could provide information to help young people choose a suitable service and prepare for their own visit. However, mental health service needs and preferences were seen as highly individual; thus reviews may deter young people from accessing services that would actually suit them. One young person suggested that a government “check mark” (ie, accreditation), indicating that services were legitimate, could be an alternative to reviews, and a parent suggested including a feature where young people and parents could post questions to be answered by a website staff member.

User accounts were the least desirable function of a service navigation website. Young people’s responses to implementing user accounts were overwhelmingly negative, due to concerns around privacy, security, safety, stigma, and limiting access to website features for people without accounts. Young people thought their peers may also associate accounts with costs like subscription fees. A small number of survey participants thought it would be helpful to have a single digital record of the services they had tried and a “wish list” of services they would like to try in the future. However, most parents and young people were uncomfortable with the idea of their mental health information being stored online and believed that this could create a barrier to people using a service navigation website. Young people
emphasized the need to be able to access a service navigation website anonymously, particularly when living in high-risk situations where internet use was monitored, and some felt that an optional user account would overcomplicate a navigation website. Young people suggested a range of other, less invasive, ways to tailor user experiences, including features that made it easy to save pages to favorites in a browser, copy content to the clipboard, email a link to service contact information, and save records of service contact information into another application (eg, into the Notes or Photos app on a phone).

A record of interactions makes me feel insecure, because I know the website is keeping data on me. This is not just “not helpful,” but distinctly unhelpful. [Young person, survey participant]

I would be concerned about a portal where you sign in. While it might be good to have information all in one place, really mental health information belongs in a doctor’s office. If your child is displaying mental health problems, you don’t really want that recorded when you don’t know how that information will be used into the future. [Parent, survey participant]

Discussion

We conducted focus groups and online surveys to inform the design of a mental health services navigation website. Although the research activity was conducted with a specific region in mind, the findings may assist other organizations designing websites or apps to assist young people and parents in navigating mental health systems. Many of the themes represent common issues faced by young people in need of mental health support.

Participants emphasized the need for, and importance of, up-to-date and accurate information about local mental health services (public and private) and guidance on how to access them (including referral pathways). Participants were aware of existing lists and directories of services, but these were described as incomplete, out of date, and difficult to search. Navigation tools were seen to be most helpful if they could provide local, tailored information, including service information that could not be accessed elsewhere. Transparent information about service cost, wait times, and how to access services without parental permission was in particularly high demand. This information could support young people and parents to make informed choices about which services to select and pursue. Future research exploring the specific service information that is needed to adequately meet end-user needs and that translates into actual service contacts would be beneficial.

Similarly to previous research, participants also emphasized the need for information about referral pathways and how to navigate the mental health system [9,19]. Parents and caregivers shared that they often acted like case managers and advocates for their children, and they wanted access to information and resources that could support them. Step-by-step instructions, flow charts, and real-life stories of accessing mental health care, tailored to the local system, could all support young people and their parents on their journeys from first seeking help to accessing specialist services [19].

An effective search engine was an important aspect of the navigation tool’s design. All information within the navigation tool needs to be searchable, and users should be able to refine their search results using relevant filters. Young people and parents felt that they should be able to independently find service information more easily than they could with a web search engine and that a navigation tool should be broadly accessible, functional, easy to navigate, and tailored to its audience in content and appearance. This finding highlights again the importance of the co-design process, and end-user testing, to ensure that the tool developed is fit for purpose and fully meets the needs of those who will use it [15]. Relatedly, digital tools should also be designed in accordance with web accessibility standards to ensure that all users can effectively use them, including those with disabilities [20].

Participants also felt that it would also be helpful for the navigation tool to provide a phone, text, and/or web chat service. Specifically, participants wanted a phone line or web chat option that could help people to identify relevant and appropriate mental health services and support them to decide which service to contact first. Including this feature would reduce the burden on young people and parents to identify, research, and select services alone. Offering an option to receive a follow-up call or text from a phone line or web chat may also be beneficial.

Underlying many of these findings was the need to have control over the help-seeking process. Control over how and when to interact with content is an important concern for young people in the design of online mental health interventions [16,21]. Providing choices was important to participants in our study; both choice in how they could contact a support person through the navigation tool (eg, phone call, web chat, SMS text messaging, or email) and choice in if, how, and when they were contacted by the navigation tool or engaged with a service.

Other potential navigation tool elements, such as being able to create a user account, were not as desirable. Storing records of health information online, particularly mental health information, raised serious concerns around privacy. Generally, participants felt that people would primarily use the navigation tool anonymously and independently. This aligns with previous findings from the development of mental health interventions for young people and adults, which highlighted confidentiality, privacy, and trust in the organization delivering the service as key areas of importance for end users [21-23]. Some people felt that it could be helpful to be able to record their activity (eg, service wish lists) while using the navigation tool and/or service use in a single location, but participants emphasized that user account features should be opt-in. Such features may be off-putting for potential users.

The involvement of end users in the design of the navigation tool was highly valued by the commissioners of this research and resulted in the development of a youth navigation website called MindMap [24], which captures the key elements identified by participants. The findings of this study provided formative information for the development of MindMap, although there was additional development and testing involved to create the final website, which is beyond the scope of this paper. MindMap is an accessible web-based tool that provides a comprehensive
searchable database of local services and provides a clear description of the service, its location, and potential wait times. It is an initiative of the ACT Office of Mental Health and Wellbeing and delivered by a nongovernmental organization with strong connections into the local mental health sector, enabling frequent refreshing of service information. Users of MindMap can use the navigation tool independently or receive navigation support during the week or on weekends from members of the MindMap team via telephone, email, or web chat. Young people and parents or caregivers were involved in the iterative development and testing of MindMap to ensure that it continued to meet their needs.

There are some limitations to this research that should also be considered. First, the participants in the focus groups and survey may not have been representative of all young people and parents or caregivers in the community, and the study may have attracted people with a greater interest in mental health. Preferences for website appearance, content, and features may have varied by age and personal experience of mental health issues [16,25]. Future research would benefit from sampling more young men, and young people and parent or caregivers without mental health service experience, to ensure that the needs and preferences of all targets are adequately captured and met.

The scope of the questions in the focus groups may also have guided the discussion, placing more emphasis on the areas covered by the questions and consequently may have missed other issues. This issue was partly mitigated by providing time within focus groups for participants to identify issues not covered by the questions. The timeline of the project necessitated the use of point-by-point note-taking from focus group recordings, rather than verbatim transcripts, for data analysis. We acknowledge that this approach may have a higher risk of introducing research bias; however, this was mitigated through regular discussion of the analysis with the research team and the inclusion of qualitative survey data in the analysis. Lastly, the research was conducted with a specific region in mind, and thus not all findings may be relevant to other contexts. Overall, this study provides important insights into the navigation needs of young people and their parents or caregivers seeking mental health services and how best to support them in this process. The focus groups and surveys identified the need for tailored local information, the provision of up-to-date service details, and the opportunity for users to navigate the site independently or with support. Ensuring that young people and their parents or caregivers can access mental health services in an efficient and timely manner is essential to the longer-term health and well-being of young people. Future research assessing the effectiveness of navigation tools in meeting this goal should be strongly encouraged.

Acknowledgments

We would like to thank and acknowledge the young people, parents, and carers who participated in this research. This research was funded by the Office for Mental Health and Wellbeing, ACT Health. ALC is supported by a National Health and Medical Research Council (NHMRC) fellowship (1173146). ARM is supported by a Suicide Prevention Australia Post-Doctoral fellowship.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

ALC and PJB led the design of the study, with input from all other authors. ALC, SMM, EM, NJ, and ARM contributed to data collection. ARM and MB conducted the analyses. ALC and ARM drafted the paper. All the authors critically reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Focus group questions and prompts.

References


Abbreviations

ACT: Australian Capital Territory
LGBTIQ+: lesbian, gay, bisexual, transgender, intersex, queer
Development and Usability Testing of an mHealth Tool for Trauma-Informed Prevention of Substance Use, HIV Acquisition, and Risky Sexual Behaviors Among Adolescents: Mixed Methods Study

Carla Kmett Danielson, PhD; Angela Moreland, PhD; Austin Hahn, PhD; Devin Banks, PhD; Kenneth J Ruggiero, PhD

1Department of Psychiatry & Behavioral Science, Medical University of South Carolina, Charleston, SC, United States
2Department of Psychology, University of Missouri-Saint Louis, Saint Louis, MO, United States
3College of Nursing, Medical University of South Carolina, Charleston, SC, United States

Corresponding Author:
Carla Kmett Danielson, PhD
Department of Psychiatry & Behavioral Science
Medical University of South Carolina
67 President Street
Charleston, SC, 29425
United States
Phone: 1 8437922945
Email: danielso@musc.edu

Abstract

Background: Youth who experience traumatic events are at a substantially higher risk of engaging in substance use and sexual risk behaviors and problems (e.g., HIV acquisition) than their non–trauma-exposed counterparts. Evidence-based substance use and risky sexual behavior prevention may reduce the risk of these outcomes. Trauma-focused mental health treatment provides a window of opportunity for the implementation of such preventive work with these youth. However, overburdened clinicians face challenges in adding prevention content while implementing evidence-based treatments. Mobile health (mHealth) tools can help reduce this burden in delivering prevention curricula. Trauma-Informed Prevention for Substance Use and Risky Sexual Behavior (TIPS) is an mHealth app that was developed to aid trauma-focused cognitive behavioral therapy (TF-CBT) clinicians in the implementation of an evidence-based risk behavior prevention curriculum.

Objective: The goal of this paper is to describe the rationale for and development of the TIPS app and present the results of a mixed methods approach for the initial evaluation of its usability.

Methods: Participants included clinicians (n=11), adolescents (n=11), and caregivers (n=10) who completed qualitative interviews and an adapted version of the Website Analysis and Measurement Inventory.

Results: In total, 4 overarching themes emerged from the participants’ answers to the qualitative interview questions, demonstrating a generally positive response to the app. The themes were (1) strength of app content, (2) suggestions about app content, (3) esthetics and usability, and (4) benefits to the patient and session implementation. Clinicians, adolescents, and caregivers all agreed that the content was very relevant to adolescents and used examples and language that adolescents could relate to. All 3 groups also discussed that the content was comprehensive and addressed issues often faced by adolescents. All 3 groups of users made suggestions about the esthetics, which mostly comprised suggestions to change the font, color, or pictures within the app. Of all the groups, adolescents were most positive about the esthetics and usability of the app. Results from the Website Analysis and Measurement Inventory further illustrated the users’ favorable reaction to the TIPS app, with 100% (11/11) of clinicians, 100% (10/10) of caregivers, and most adolescents (7/11, 64%) selecting strongly agree or somewhat agree to the following statement: “This app has much that is of interest to me.” Adolescents generally found the app easier to use than did caregivers and clinicians.

Conclusions: The TIPS app shows promise as an mHealth tool for TF-CBT clinicians to integrate evidence-based substance use, risky sexual behavior, and HIV prevention during treatment. Future research, including a randomized controlled trial comparing...
TF-CBT implementation with and without the inclusion of the app, is necessary to evaluate the feasibility and efficacy of the app in reducing the risk of substance use and risky sexual behavior among trauma-exposed adolescents.

**Trial Registration:** ClinicalTrials.gov NCT03710720; https://clinicaltrials.gov/study/NCT03710720

(JMIR Form Res 2024;8:e52835) doi:10.2196/52835

**KEYWORDS**
traumatic stress; prevention; substance use; HIV; qualitative methods; adolescents; mobile phone

**Introduction**

**Background**

Children and adolescents who experience traumatic events are at a substantially higher risk of engaging in sexual risk behaviors and substance use than their non–trauma-exposed counterparts [1,2]. Early traumatic experiences also have long-term effects on behavioral health as youth exposed to trauma are more likely than their nonexposed peers to develop substance use disorders [3,4] and experience unexpected pregnancy and sexually transmitted infections (STIs) [5,6], including increased risk of HIV acquisition [7,8]. Models addressing the link between trauma exposure and risk behavior have suggested that, although these behaviors are multidetermined, trauma exposure plays an important role by manifesting impairments in affect regulation, impulse control, identity development, and socialization [2,9].

Given the strong relationship between trauma exposure and problematic behavioral health outcomes, preventing morbidity related to substance use and sexual risk behavior may be best addressed from a trauma-informed lens. That is, an important prevention approach may involve addressing risk behavior in the context of trauma-focused assessment and treatment. However, data suggest that clinicians are often reluctant to integrate interventions related to substance use in the context of trauma treatment [10] because of clinical, systemic, and training barriers and limited resources (including limited support) relative to caseload demands [10-13]. Given the need for substance use and risky sexual behavior prevention for trauma-exposed adolescents—and to combat these barriers and demands for time and resources—clinicians may benefit from a structured tool to help efficiently implement evidence-based prevention strategies for risk behaviors common among trauma-exposed youth [14]. Such a tool may be best delivered via a mobile health (mHealth) approach, which can reduce the need for extensive clinician training in multiple new prevention intervention curricula, augment the formal training they do receive, and enhance or extend the effectiveness of the traditional clinical encounter [15,16]. As such, clinicians would be significantly more likely to deliver substance use prevention if a ready-made mHealth tool were available that helped them deliver it efficiently and accurately. This study describes the development and feasibility evaluation of an mHealth app designed to supplement trauma treatment among adolescents, targeting prevention of substance use, sexual risk behavior, and associated health consequences (eg, HIV).

**Adolescent Trauma Treatment as an Opportunity for Prevention**

The gold-standard treatment for addressing adolescent trauma is trauma-focused cognitive behavioral therapy (TF-CBT) [17,18]. With >20 completed randomized controlled trials (RCTs)—including international studies—supporting its effectiveness in addressing a range of mental health problems and improving functioning among trauma-exposed youth, TF-CBT has received the highest ranking for empirical support from professional organizations and federal agencies [19-38]. TF-CBT has achieved widespread dissemination through numerous implementation efforts, including comprehensive training programs (eg, community-based learned collaboratives), educational materials, and collaborations with mental health organizations, ensuring its accessibility and use by therapists and clinicians across various regions and populations. In fact, the widespread utility of TF-CBT is exemplified by the staggering number of clinicians who have accessed TF-CBT Web 1.0 and 2.0 (original and updated versions of the TF-CBT web-based learning course); as of August 31, 2023, a total of 496,061 clinicians worldwide are registered users of the web-based training in the TF-CBT model, and 305,120 clinicians have completed the training [39]. Modules of TF-CBT, including psychoeducation and enhancing safety, may include general psychoeducation about sex and sexual revictimization risk reduction; however, information and skill development specific to HIV and STIs, pregnancy, and healthy dating and sexual decision-making are not detailed in the model manual or training. Similarly, skill development for preventing substance use problems that extends beyond psychoeducation and helps translate this knowledge into skills is not formally or typically incorporated into TF-CBT, limiting the potential impact of these psychoeducation modules. In other words, these clinicians do not systematically receive the training and support needed to feel confident in their delivery of psychoeducation content or in the ways to translate this education into skills (eg, realistic refusal skills) [10].

Although clinicians can informally incorporate risk behavior topics into their implementation, the data suggest that this is uncommon. A national survey of mental health clinicians found that providers feel ill-equipped to address topics of substance use and sexual risk behavior when treating trauma-exposed adolescents even when trained in trauma-informed treatment models [10]. Most clinicians did not report receiving formal training for addressing substance use disorder (54%) or sexual risk behavior (67%), suggesting that this reluctance may be related to deficits in training on these topics [10]. These data reflect a global lack of training in evidence-based practice for adolescent prevention, which results in limited translation and...
accessibility despite the availability of numerous efficacious preventive interventions [40]. In summary, although effective sexual risk and substance use prevention interventions are available, there is a significant gap in the implementation of these interventions even among those at high risk of such behaviors because of trauma exposure.

**mHealth as a Viable Approach to Address Prevention**

The availability of mobile technology has increased dramatically over the past decade, with 85% of Americans reporting smartphone ownership in 2022 compared with 35% in 2011 [41,42] and 95% of American adolescents reporting smartphone access [43,44]. As mobile technology has rapidly developed, so have health care approaches that leverage mHealth—the use and development of mobile technology, including mobile apps, to improve health care [16,45]. Within mental and behavioral health care, mHealth interventions are diverse, targeting different stages of treatment from education and engagement to the maintenance of treatment gains [16]. mHealth administrations are also diverse and may be stand-alone, client-led interventions or supplements to traditional clinician-facing treatments [16].

Most of the extant mHealth approaches to behavioral health among adolescents are stand-alone interventions rather than supplements to face-to-face treatment [46]. Several stand-alone treatments have shown strong feasibility and efficacy for both primary and secondary prevention of substance use and sexual risk behavior [47-50]. However, most stand-alone behavioral interventions for adolescents lack a theoretical framework in their design and show inconsistent efficacy [46,51]. Furthermore, although data suggest that digital mental health tools delivered in real-world contexts are more likely to be accessed than professional services, they are less likely to be delivered with a sufficient therapeutic dose [52]. These limitations are compounded by clinician- and patient-reported barriers to stand-alone treatments, including the lack of personalization [51,53], lack of privacy and security with regard to sensitive behavioral topics [53,54], limited follow-up [51], and attrition and low completion rates [54,55]. Researchers have suggested that these barriers could be overcome with more support and involvement from clinicians [53]. Accordingly, behavioral health treatments that blend face-to-face and technology-based approaches have been found to save clinician time, demonstrate lower dropout rates, and lead to better treatment outcomes among adolescents and young adults [56,57].

With regard to trauma treatment, 3 popular mHealth apps have emerged for use with adults: Posttraumatic Stress Disorder (PTSD) Coach, Cognitive Processing Therapy Coach, and Prolonged Exposure Coach [58]. The most widely used of these is PTSD Coach, a stand-alone treatment that suffers from limitations similar to those mentioned previously [58], including high attrition and inconsistent efficacy [59]. Conversely, Cognitive Processing Therapy Coach and Prolonged Exposure Coach are supplementary to existing evidence-based treatments for PTSD. Although these apps have not been evaluated for efficacy via RCTs, clinician perceptions of such apps have been favorable, particularly regarding their relative advantage over exclusively face-to-face practices and compatibility with clinicians’ needs [60]. In addition, supplementary mHealth approaches have demonstrated effectiveness in the treatment of comorbid trauma conditions (eg, panic disorder, anxiety, and depression) [61], including comorbid PTSD and substance use [62].

Despite evidence of the feasibility and effectiveness of mHealth interventions in trauma treatment among adults, mHealth approaches to trauma treatment among adolescents are limited. In total, 3 stand-alone prevention approaches for adolescents exposed to acute trauma have demonstrated effectiveness in reducing persistent trauma symptoms, depressive symptoms, and behavioral problems [63-65]. Although no studies have examined the effectiveness of supplementary mHealth approaches for adolescent trauma treatment, the perspectives of trauma-focused clinicians indicate that these approaches would be feasible and useful. For example, Orengo-Aguayo et al [66] found that 96% to 100% of surveyed TF-CBT providers reported that an mHealth supplement would be helpful for enhancing TF-CBT components, extending coping skill development, and improving out-of-session practice among adolescents and families. Clinicians, patients, and families have also responded favorably to pilot versions of a supplemental Apple iPad–based app designed to improve patient engagement and provider fidelity in-session during TF-CBT [67,68]. Thus, mHealth supplements to adolescent trauma treatment may be a feasible and acceptable way to extend and improve behavioral health prevention among adolescents.

**This Study**

The purpose of this paper is to report on the development and perceived usability of Trauma-Informed Prevention for Substance Use and Risky Sexual Behavior (TIPS), a novel mHealth tablet-based app designed to supplement trauma treatment targeting the prevention of adolescent risk behavior (trial registration: ClinicalTrials.gov NCT03710720). We used a mixed methods approach based on qualitative interviews and quantitative ratings to assess the usability of the TIPS app with clinicians, adolescent patients, and their caregivers engaged in TF-CBT in a community-based outpatient clinic.

**Methods**

**TIPS App Structure**

The first author, a national trainer in the TF-CBT model, led a small team of TF-CBT clinicians and trainees in the development of the content of the TIPS mHealth app. This included 7 total topics (Figure 1) designed to be used as psychoeducational tools for TF-CBT clinicians to implement with adolescents and caregivers throughout the TF-CBT treatment process. The tool helps clinicians assess their clients’ and caregivers’ current knowledge and comfortability surrounding topics related to risky sexual behaviors, STIs, drug use, and healthy relationships. Thus, there are 3 intended users of the app: clinicians, adolescents, and caregivers. Each user sets up a unique log-in, and the content displayed on the app is tailored to the user type. For example, Family Check-Up example videos demonstrating parenting skills are displayed for the caregiver but not for the youth, and a sexting decision-making activity is displayed for the youth but not for...
the caregivers; clinicians view both caregiver and youth content but also have unique introduction videos for each app topic tailored to a clinician audience and have additional drop-down menus on the app, such as suggested homework that the clinician can assign.

Figure 1. Trauma-Informed Prevention for Substance Use and Risky Sexual Behavior (TIPS) home screen (after log-in) listing the 7 topics and activities.

In particular, the app was designed to map onto the psychoeducation, parenting, and enhancing safety treatment components of TF-CBT. Each topic of the app is introduced by a short video that explains the aims of the section as well as the directions needed to successfully complete the section. After the completion of each component activity (briefly described in this section), users are presented with feedback relevant to their performance on the current activity as well as resources (links to websites) to help further their education on the topic. Specifically, the What Do You Know? component is formatted as a multiple-choice trivia game that serves as a psychoeducational tool to educate trauma-exposed teenagers and their caregivers on the effects of drugs and alcohol. The Myth Busters component educates users on the common facts and myths about HIV and other STIs. This section was designed as a drag-and-drop game and includes a video that demonstrates how to perform an at-home HIV test kit. The inclusion of the video aimed to reduce negative connotations that may currently exist related to getting tested for HIV (eg, scary, involving needles, and time-consuming). The Hot or Not component was designed as a psychoeducational tool to aid adolescents in recognizing unhealthy behaviors within romantic relationships. This section comprises engaging videos, multiple-choice trivia, and feedback that is presented after response submission. A main goal was to highlight warning signs that may precede more obvious unhealthy behavior as trauma-exposed teenagers may be more likely to stay in unhealthy romantic relationships. Make a Play helps guide adolescents through tough yet common situations that teenagers may experience. This component consists of 5 activities that educate users on choices involving sending nude pictures, having unprotected sex, consuming alcohol, and being offered different types of drugs. These activities are presented as choose-your-own-adventure screenplays that allow adolescents to make different choices for each situation that is presented and view the results for their different decisions. Figure 2 illustrates an adolescent presented with the choice to accept or decline the offer of marijuana by a peer. The Family topic of the app presents the empirically supported Family Check-Up [69] in an engaging way, educating caregivers on the different components involved in positive parenting (eg, communication, encouragement, and supervision). This section consists of videos, text information, and quizzes. Cyber Life was created to aid in the education regarding safe web behaviors and choices. This component presents users with possible situations that they may encounter on the web and the choices they can make in a multiple-choice format. Each choice is followed by feedback on its positive aspects and the risks involved. The Vision 25 topic is to be used as a guide in goal setting. This section helps users think about their current goals, what choices they can make to help them successfully accomplish their goals, and what choices they can make that may result in their goals becoming harder to reach. This section presents users with a road map (Figure 3) where they can choose different ages at which to set goals, providing guidance for both short- and long-term goals.
The TIPS platform was developed using AppBuilder, a content management system developed by our digital health team under our institution’s Clinical and Translational Science Award. AppBuilder includes a wide array of design templates and features to ensure that native iOS and Android apps can be created by researchers, program staff, and other team members who have no formal coding experience. Investigators and innovators use it to wireframe, pilot test, change and add content, and launch and evaluate mHealth apps without the extensive involvement of a technical development team. Mobile app developers often become involved in AppBuilder-based initiatives only on a strategic, limited basis, consistent with our goal of significantly shortening the timeline and costs associated with building mHealth apps.

Usability Testing Overview
Usability testing is incorporated into technology development to improve user experience by measuring whether the user can successfully and effectively use the tool. Usability testing incorporating both qualitative and quantitative methods can also help identify barriers to task completion and examine areas that take the user off topic, create confusion, or decrease satisfaction [70]. The purpose of usability testing in this study was to obtain objective metrics and refine the final TIPS product before formal efficacy testing.

Ethical Considerations
All procedures were approved by the Medical University of South Carolina Institutional Review Board (Pro00041527). Participants (ie, clinicians, caregivers, and teenagers) were given an information sheet that detailed the purpose of the research; procedures; risks, discomforts and benefits; costs and compensation; alternatives; and confidentiality. Participants provided verbal consent in lieu of written consent. Participant data were deidentified to preserve confidentiality. In addition,
participants were compensated for their participation with US $30 in the form of a gift card, cash, or money order.

**Participants**

Participants included mental health clinicians (11/32, 34%; 26/32, 82% female), trauma-exposed adolescents (11/32, 34%; 26/32, 82% female), and caregivers of adolescents (10/32, 31%; 25/32, 80% female). More specifically, with regard to inclusion criteria, clinicians were master’s or doctoral-level mental health providers in the area local to the study who were fully trained and experienced in TF-CBT and carried active child trauma caseloads. Adolescents were aged between 13 and 18 years, had experienced at least one traumatic event, were in the process of completing or had recently completed TF-CBT, and assented to participate (with caregiver consent). To be included, caregivers needed to have served for at least the previous 2 months in the role of primary caregiver of a teenager in treatment for PTSD who was in the process of completing or had recently completed TF-CBT. Among the clinicians, 72% (8/11) were aged between 25 and 34 years, whereas 27% (3/11) were aged between 35 and 44 years. The adolescents ranged in age from 13 to 18 (mean 15.25, SD 1.90) years, with most (9/11, 82%) falling between the ages of 13 and 17 years. Finally, caregivers ranged in age from 34 to 53 (mean 44.2, SD 6.65) years.

**Procedures**

Participants (clinicians, adolescents, and caregivers) were recruited from community-based mental health centers that specialized in the treatment of traumatic stress and served multiple counties (including urban and rural areas) in a large city in the Southeast United States. There were 2 primary methods through which participants were recruited: (1) flyers were posted in and around the clinics with contact information for the study project coordinator and (2) potential participants were informed by clinic staff that they may qualify for a study and asked whether they would be interested in learning more. When potential participants responded positively, the study coordinator contacted them and provided a full description of the study, screened them to ensure that the inclusion criteria were met (see the previous section), and consented them to participate in a qualitative interview regarding the TIPS app and completion of the Website Analysis and Measurement Inventory (WAMMI; see the following section).

**Measures**

The semistructured qualitative interview consisted of a study team member providing the participant with an iPad loaded with the TIPS app and walking the participant through each section. Following each section of the app, the interviewer asked several open-ended questions (eg, “First, tell me, what are you thinking when you look at this page?” “What do you like about this activity?” “What don’t you like about this activity?” and “How can we make this more interesting to teens?”). Follow-up probes were used to clarify the information provided whenever necessary.

The WAMMI [71] is a standardized 20-item assessment measure that captures users’ personal opinions on a given website’s ease of use. The items were slightly revised to refer to the application rather than the website. In the measure, users are asked to rate various aspects of their experience with the app (eg, content, navigation, and design) on a 5-point Likert scale from strongly agree to strongly disagree. Items are then scored to produce 5 subscales measuring the app’s attractiveness (level of visual interest of the app in terms of both function and information provision), control (app navigation ease), learning (users’ ability to easily understand the content of the app and learn what they expect to learn), helpfulness (usefulness and expected content and structure of the app), and efficiency (users’ ability to quickly find and do what interests them on the app). The WAMMI was developed using latent variable analysis, has high reliability, and reports standardized scores (eg, 50=average; 100=perfect) for the 5 aforementioned themes based on a reference database [72].

**Data Analysis Plan**

**Quantitative Approach**

All descriptive analyses were performed using Stata (version 17; StataCorp LLC) [73]. The data were screened for outliers and impossible values. Group differences in the WAMMI items were compared using 2-tailed paired t tests.

**Qualitative Approach**

Data analysis consisted of a qualitative content analysis [74] informed by grounded theory [75], which is used to explore participants’ unique perspectives via the identification of themes and patterns that naturally emerge from the data and the systematic classification of these themes. Specifically, a 3-step inductive approach was used, which involves collecting and analyzing data without preconceived categories or theories. To analyze using this approach, each participant’s interview responses (ie, raw data) were carefully examined to develop a comprehensive codebook to capture all possible themes emerging from the data. The codebook was then used by 2 independent coders to code and analyze each participant’s responses to the interview questions [74,76]. Coders were able to apply more than one code to the participant responses if applicable. The interrater reliability for the double-coded interview responses was 86% and ranged from 82% to 93%. Interrater discrepancies were discussed and resolved by the 2 independent coders. Finally, themes were refined, merged, or divided into subthemes via collaborative discussions in multiple in-person meetings until a comprehensive codebook was developed. The NVivo software (version 11.1; Lumivero) was used for data management and analysis. The interviews were approximately 45 minutes in length and were audio recorded and transcribed.

**Results**

**Technology and App Use Descriptive Information**

Clinicians answered some descriptive questions about comfort using smartphones and apps and the benefits and drawbacks of using websites or apps in treatment. All clinicians (10/10, 100%) reported (1) being comfortable using smartphones and tablets and (2) that the primary benefit of using websites and mHealth app tools is that they engage teenagers in treatment and that they are readily accessible. Other benefits reported by clinicians
included that apps are free and that teenagers feel comfortable using them. Drawbacks reported by clinicians included that it is difficult to get teenagers to stop using apps and that not all homes have Wi-Fi access.

Adolescents and caregivers were asked about their use of tablets and cell phones. Most adolescents (8/11, 73%) and caregivers (8/10, 80%) reported personally owning a tablet, and all adolescents (11/11, 100%) and caregivers (10/10, 100%) reported owning a cell phone. The primary uses of tablets by adolescents included accessing social media, playing games, watching television shows or movies, SMS text messaging, and completing schoolwork. Adolescents reported that they primarily used their cell phones for talking to friends, SMS text messaging, accessing social media, playing games, and listening to music. Caregivers reported that they primarily used their tablets for playing games, paying bills, surfing the internet, watching television shows and movies, school and work, accessing social media, and checking email. Caregivers primarily used their cell phones for talking to people, SMS text messaging, accessing social media, and surfing the internet.

### Qualitative Results

#### Overview

Through the individual interviews, valuable information about the usability and perceived effectiveness of TIPS was obtained from clinicians, adolescents, and caregivers. Four overarching themes, each with its own subthemes, emerged from the participants’ answers to the interview questions: (1) strength of app content, (2) suggestions about app content, (3) esthetics and usability, and (4) benefits to the patient and session implementation. Table 1 shows the app themes that emerged. Each is described in greater detail in the following sections, with representative quotes provided throughout for illustrative purposes.

#### Table 1. Trauma-Informed Prevention for Substance Use and Risky Sexual Behavior (TIPS) app themes and percentages yielded from the qualitative interviews

<table>
<thead>
<tr>
<th>Theme</th>
<th>Responses, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strengths of app content</strong></td>
<td></td>
</tr>
<tr>
<td>Comments regarding appropriateness of content for adolescents</td>
<td>27 (13.9)</td>
</tr>
<tr>
<td>Content of modules</td>
<td>10 (5.2)</td>
</tr>
<tr>
<td>Pictures or videos</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>Content is entertaining or would keep adolescents’ attention</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td><strong>Changes in or suggestions for app content</strong></td>
<td></td>
</tr>
<tr>
<td>Make content more engaging</td>
<td>17 (8.8)</td>
</tr>
<tr>
<td>Revise the video or picture content</td>
<td>9 (4.6)</td>
</tr>
<tr>
<td>Suggestions for content topics to add</td>
<td>20 (10.3)</td>
</tr>
<tr>
<td>Revise the language to make it more relatable to teenagers</td>
<td>22 (11.3)</td>
</tr>
<tr>
<td>Allow content to be more individualized</td>
<td>15 (7.7)</td>
</tr>
<tr>
<td><strong>Esthetics and usability</strong></td>
<td></td>
</tr>
<tr>
<td>Strengths regarding esthetics of the app</td>
<td>8 (4.1)</td>
</tr>
<tr>
<td>Suggestions for esthetics of the app</td>
<td>17 (8.8)</td>
</tr>
<tr>
<td>Dislikes regarding esthetics of the app</td>
<td>5 (2.6)</td>
</tr>
<tr>
<td><strong>Benefits to the patient and session implementation</strong></td>
<td></td>
</tr>
<tr>
<td>Increases comfort of the adolescent in the session</td>
<td>5 (2.6)</td>
</tr>
<tr>
<td>Would be more likely to use the TIPS app</td>
<td>2 (1)</td>
</tr>
</tbody>
</table>

*Themes that emerged from the responses of all users during the qualitative interviews (32 participants and 194 responses).

#### Theme 1: Strengths in the App Content

A total of 62% (20/32) of the participants discussed the strengths of the content of the TIPS app. Most strengths mentioned within this theme included comments about the appropriateness of the content for adolescents (21/32, 66%), the content of the modules (17/32, 53%), pictures or videos (13/32, 41%), or that the content was entertaining and would keep an adolescent’s attention (11/32, 34%). Other comments centered on language content, diversity of content, positive framing of information, or that the content differed by activity. Clinicians, adolescents, and caregivers all agreed that the content was very relevant to adolescents and used examples and language that adolescents could relate to. All 3 groups also discussed that the content was comprehensive and addressed issues often faced by adolescents. For example, clinicians commented the following:

*I really like the feedback sessions when you give the examples. They’re very detailed, but not overly detailed.*

*I think the language is really accessible to most kids.*
Examples of strengths stated by adolescents include the following:

The videos are funny. I definitely like those.
I actually think that the text itself, that was pretty cool how y’all made it like the iPhone with all the emojis and stuff. And it is kind of relatable to people my age, at least, text. And the responses to it were...like sending the cat picture, that’s definitely something people would do if they don’t feel comfortable and they’re trying to make it funny.

Strengths mentioned by caregivers included the following:

I think the Let’s talk about Sext [video] needs to come as soon as you put the phone in their hand. Plain and simple, right off the rail. It’s not like there’s going to be a certain age, it’s like having the talk about the birds and the bees.

Man, I liked the points they’re making. You know, I think they’re absolutely making the right points with, you know, being present, you know, when your kids are around. Knowing your kids’ friends, knowing where they are, surprise phone calls. I like all that stuff.

**Theme 2: Changes in or Suggestions for App Content**

A total of 59% (19/32) of the participants discussed changes or suggestions regarding the TIPS app content. Most comments within this theme included making content more engaging (22/32, 69%), revising the video or picture content (22/32, 69%), suggestions for content topics to add (17/32, 53%), revising the language to make it more relatable to teenagers (15/32, 47%), and allowing the content to be more individualized (8/32, 25%). Other comments included language suggestions, adding more outside resources, increasing cultural sensitivity, clarifying content, improving instructions, and removing potentially triggering content. Although the specific suggestions differed among the clinicians, adolescents, and caregivers, the feedback overlapped in that adding more interaction and individualizing some of the content would help improve the app. Some examples of comments made by clinicians include the following:

I’ve seen a lot of kids get through school that can’t read, so try to look at all factors. Maybe some more visual aids along with the words.

Some statements made by adolescents included the following:

Add a game or something. I don’t know what type of game, but a game that you can play for the answer.
Maybe even include something about rape in a relationship, because a lot of people that I’ve talked to, they don’t think that you can be raped in a relationship.

Comments made by caregivers included the following:

You know, make it a little bit more interesting. If it could be made more interactive.
A lot of kids don’t have parents. They’re already living with a family member, or in foster care. Another line to put in there somewhere is, talk to someone you trust...doesn’t have to be a parent.

**Theme 3: Esthetics and Usability**

A total of 25% (8/32) of the participants discussed the esthetics and usability of the TIPS app. The most common comments within this theme involved strengths regarding the esthetics of the app (15/32, 47%), suggestions for the esthetics of the app (13/32, 41%), and dislikes regarding the esthetics of the app (9/32, 28%). Participants also made a few comments about usability, including suggestions and strengths regarding the functionality of the app. All 3 of the groups (clinicians, adolescents, and caregivers) made suggestions about the esthetics, which mostly comprised suggestions to change the font, color, or pictures within the app. Of all the groups, adolescents were the most positive about the esthetics and usability of the app. Some specific comments made by clinicians included the following:

Maybe use highlight and change the color or something to make it clear. I think that the progress bar at the bottom is so simple. It stays out of the way, so you’re not sacrificing real estate.
[I like] the progress bar on the bottom. It’s a nice touch because then I don’t have to keep wondering how many more questions. I can see that I’m almost done, so that helps me not get frustrated, especially for a teen.

Some examples of statements made by adolescents include the following:

The activities are very easy to operate.
I think it’s a good font size, I wear glasses, and I didn’t have to squinch my eyes to see. I also like how the page is light blue, and it’s dark blue for the words to stick out. So you’ll be able to look at it. That’s one thing I like about it.

Specific comments made by caregivers included the following:

I like the subtitles on the bottom, because I feel like you sometimes get lost in the words, so I like that they’re there.
I like the bright colors. The bright colors catch my eye.

**Theme 4: Benefits to the Patient and Session Implementation**

A total of 19% (6/32) of the participants discussed how the TIPS app benefits the patient or made comments about session implementation. Most comments within this theme included that the TIPS app increases the comfort of the adolescent in the session (10/32, 31%) or that clinicians would be more likely to use the TIPS app with particular patients (10/32, 31%), such as female individuals or sexually active teenagers. Other comments in this theme included that the app makes the session more interactive, increases engagement for teenagers, is easy to incorporate into the session, and allows the clinician to modify or select content based on patient needs. Some examples of comments made by clinicians include the following:
This is good for me, too, the way I’m able to give feedback to the patient, or at least talk about it, cuz once it’s there, okay, I get it, and I can explain it to them, so I really like it.

Comments made by adolescents included the following:

The app makes it easier because it’s just awkward to talk about these things.

Some statements from caregivers included the following:

I think for this exercise as far as getting them to wrap their brain around where they really have established rapport with you, it’s probably a nice way of doing that where they don’t really have to look you in the eye and tell you about what they want.

Quantitative Results

The results of the WAMMI across the 3 groups are shown in Table 2. Responses across the groups indicated that all participant groups viewed the app favorably. In general, clinicians tended to report the most critical responses to the app. Overall, adolescents found the app (relatively) easier to use and understand compared with clinicians and caregivers. Clinicians had a greater propensity to report the app having annoying features compared with caregivers and adolescents. All clinicians (11/11, 100%) and caregivers (10/10, 100%) and 64% (7/11) of adolescents selected somewhat agree or strongly agree for the following statement: “This app has much that is of interest to me.” All adolescents (11/11, 100%) somewhat agreed or strongly agreed with the following statements: “I can quickly find what I want on this app,” “Using this app for the first time is easy,” and “Everything on this app is easy to understand.”
Table 2. Website Analysis and Measurement Inventory (WAMMI) responses regarding the Trauma-Informed Prevention for Substance Use and Risky Sexual Behavior app by group.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Clinicians (n=11)</th>
<th>Caregivers (n=10)</th>
<th>Adolescents (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>This app has much that is of interest to me.</td>
<td>1.73 (1.19)</td>
<td>1.20 (0.42)</td>
<td>2.36 (0.81)</td>
</tr>
<tr>
<td>It is difficult to navigate this app.</td>
<td>2.91 (1.04)</td>
<td>3.90 (1.79)</td>
<td>4.64 (0.92)</td>
</tr>
<tr>
<td>I can quickly find what I want on this app.</td>
<td>2.36 (1.03)</td>
<td>1.40 (0.70)</td>
<td>1.45 (0.52)</td>
</tr>
<tr>
<td>This app seems logical to me.</td>
<td>1.55 (1.21)</td>
<td>1.50 (1.27)</td>
<td>1.36 (0.67)</td>
</tr>
<tr>
<td>This app needs more introductory explanations.</td>
<td>3.45 (0.93)</td>
<td>3.80 (1.40)</td>
<td>3.82 (1.54)</td>
</tr>
<tr>
<td>This app is very attractive.</td>
<td>2.55 (1.37)</td>
<td>1.90 (1.29)</td>
<td>2.36 (1.03)</td>
</tr>
<tr>
<td>I feel in control when I’m using this app.</td>
<td>2.09 (1.04)</td>
<td>1.50 (0.97)</td>
<td>1.73 (0.79)</td>
</tr>
<tr>
<td>This app is too slow.</td>
<td>4.0 (1.0)</td>
<td>3.90 (1.60)</td>
<td>4.18 (0.98)</td>
</tr>
<tr>
<td>This app helps me find what I am looking for.</td>
<td>2.09 (0.94)</td>
<td>1.70 (1.06)</td>
<td>1.73 (0.65)</td>
</tr>
<tr>
<td>Learning to find my way around this app is a problem.</td>
<td>3.45 (1.04)</td>
<td>4.30 (1.34)</td>
<td>4.73 (0.47)</td>
</tr>
<tr>
<td>I don’t like using this app.</td>
<td>4.09 (1.58)</td>
<td>4.80 (0.42)</td>
<td>4.09 (1.04)</td>
</tr>
<tr>
<td>I feel efficient when I’m using this app.</td>
<td>2.27 (1.19)</td>
<td>1.60 (0.97)</td>
<td>2.09 (1.14)</td>
</tr>
<tr>
<td>It is difficult to tell if this app has what I want.</td>
<td>3.18 (1.47)</td>
<td>3.90 (1.66)</td>
<td>3.72 (1.19)</td>
</tr>
<tr>
<td>Using this app for the first time is easy.</td>
<td>2.64 (1.21)</td>
<td>1.50 (1.27)</td>
<td>1.45 (0.52)</td>
</tr>
<tr>
<td>This app has some annoying features.</td>
<td>2.73 (1.49)</td>
<td>4.40 (1.07)</td>
<td>4.18 (1.17)</td>
</tr>
<tr>
<td>Remembering where I am on this app is difficult.</td>
<td>3.18 (1.08)</td>
<td>3.70 (1.70)</td>
<td>4.82 (0.40)</td>
</tr>
<tr>
<td>Using this app is a waste of time.</td>
<td>4.45 (0.93)</td>
<td>4.90 (0.32)</td>
<td>4.64 (0.92)</td>
</tr>
<tr>
<td>I get what I expect when I click on things on this app.</td>
<td>2.00 (1.10)</td>
<td>1.80 (1.48)</td>
<td>1.55 (0.52)</td>
</tr>
<tr>
<td>Everything on this app is easy to understand.</td>
<td>2.45 (1.13)</td>
<td>1.30 (0.48)</td>
<td>1.36 (0.50)</td>
</tr>
</tbody>
</table>

---

Discussion

Principal Findings

As of October 2021, a national state of emergency in child mental health has been jointly declared by the American Academy of Pediatrics, American Academy of Child and Adolescent Psychiatry, and Children’s Hospital Association because of the rising rates of behavioral health problems coupled with a limited and overburdened mental health workforce [52]. To address this public mental health crisis, it is critical to enlist a broad and creative range of approaches, including those that leverage mHealth tools, to implement empirically supported content and interventions that target the significant drivers of this state of emergency. Also critical to this crisis is the need for more robust prevention of substance use problems (including opioid overdose and opioid use disorders) and HIV and other STI acquisition (particularly among young people at the highest
risk for new HIV acquisition) [77,78], which is signaled by every public health indicator. The goal of this paper was to describe the development and usability testing of an mHealth app tool that collectively targets the aforementioned problems (ie, high prevalence of trauma-related mental health problems, substance use disorder, and opioid overdose and new HIV acquisition risk, as well as limited resources and opportunities dedicated to implementing substance use prevention and training clinicians in doing so).

The primary overall finding from the usability testing of the app leveraging a mixed methods approach was that the TIPS app was perceived by all 3 user types to be a highly usable mHealth tool to be implemented during the course of TF-CBT. Qualitative data collected from the usability testing of the app yielded positive feedback from clinicians, adolescents, and caregivers, and the quantitative data (ie, responses to the WAMMI) concurred with the qualitative findings (eg, 27/32, 84% of all users agreed or strongly agreed that “The app has much of interest to me”). However, more generally, the results of this study provide valuable insights into the use of technology and apps in adolescent mental health treatment. Clinician responses illustrated that they generally felt comfortable using smartphones and tablets and believed that these mHealth tools yielded benefits of engagement and accessibility for their adolescent clients. Moreover, the fact that all clinicians reported comfort with using these devices suggests a high level of digital readiness among this workforce in need of resources and support, which is a shared sentiment among other clinicians [79]. Despite this positive feedback, clinicians also identified potential drawbacks related to apps and mHealth tools, including the challenge of teenagers potentially overusing or being distracted by apps and issues related to limited wireless internet access in some homes. Indeed, equitable access to technology is essential when developing, evaluating, and implementing mHealth tools as augmentations to mental health interventions.

Specific to the interview responses, 4 overarching themes related to the TIPS app’s usefulness and effectiveness emerged. The first theme, strengths of the app content, highlights the positive aspects of the app, such as content tailored well to an adolescent audience, modules that are viewed as engaging, and relatable language. These strengths are crucial for maintaining adolescent clients’ attention and fostering meaningful interactions during trauma treatment, particularly when talking about what may be perceived as sensitive topics (eg, sexual decision-making). The second theme, changes in or suggestions for app content, points to the need for interactivity, revisions in video and picture content, and the inclusion of important topics such as consent and rape in relationships. Echoing findings of previous studies, these suggestions underscore the importance of user input and continuous improvement and adaptation of app content to meet evolving needs [80]. This also highlights the importance of using tools that allow for efficient and inexpensive minor edits (eg, to language and images) to mHealth apps when possible, as was done using the AppBuilder platform for the TIPS app. The third theme, esthetics and usability, emphasizes the significance of the app’s design and functionality, with users offering both praise and suggestions for improvement. Beyond further underscoring the importance of having the capacity to revise content within an app as part of the iterative app development process, these praise interview responses also help inform implementation strategies, indicating what end users like most and find most engaging about the app. For example, regarding TIPS, the Make a Play activity emerged as a favorite (eg, the opioid pill activity and Let’s talk about Sex!), and it may be an activity to highlight when first explaining the app to a clinician or adolescent client (eg, engaging). Finally, the fourth theme, benefits to the patient and session implementation, highlights the app’s potential to enhance adolescent and caregiver comfort during sessions and improve engagement, particularly for sensitive topics. It also suggests that clinicians are more likely to use the app with specific patient demographics, such as sexually active teenagers.

Regarding one of the clinicians’ comments that it can be difficult to get teenagers to stop using apps, it is important to note that a detailed implementation manual for the TIPS app is provided to clinicians when they are trained in how to use the app. Specifically, they are guided on how to structure the time spent on activities on the app, which occurs in the context of a TF-CBT treatment session. Thus, TF-CBT clinicians are able to contain adolescents’ use of the TIPS app in sessions.

The quantitative results, presented in Table 1, reinforce the positive reception of the TIPS app across all participant groups. Adolescents in particular found the app less difficult to use and more understandable compared with clinicians and caregivers, but clinicians expressed some critical feedback, including minor annoyance with certain features. Notably, all clinicians and caregivers, along with most adolescents, expressed a strong interest in the app’s content. In addition, adolescents found the app easy to navigate and understand, suggesting that user-friendliness is a key factor in their engagement with digital mHealth tools [81]. Overall, these quantitative findings align with the qualitative feedback, highlighting the promising utility of the TIPS app for clinician implementation among adolescents and caregivers receiving TF-CBT while also emphasizing areas for improvement.

A strength of the TIPS tool is that it offers a platform that is telehealth compatible. The past several years have underscored the value of ensuring access to mental health treatment and risk behavior prevention through telehealth strategies. In addition to the benefits of breaking down geographical barriers, the integration of mHealth tools such as TIPS with telehealth helps promote confidential and convenient interactions for patients with mental health clinicians, fostering a sense of privacy and comfort that is crucial for therapy [82,83]. This may be particularly helpful when addressing sensitive topics such as trauma, substance use, and HIV. Ultimately, the synergy between mHealth tools and telehealth in behavioral care represents a transformative shift toward prevention-focused holistic and inclusive mental well-being support.

Limitations
The primary limitation of this study is that it is limited in scope to usability testing with a small sample of clinicians, adolescents, and caregivers. To establish the feasibility and efficacy of the app in reducing the risk of substance use and risky sexual behavior, a fully powered RCT is necessary to compare TF-CBT...
implementation with and without the inclusion of the app, including assessment time points that follow youth and their caregivers over time. Although this study informs possible revisions to the app and suggests that clinicians, adolescents, and caregivers will respond positively to its inclusion in trauma-focused treatment, the efficacy trial will ultimately reveal whether indeed the app is able to make a dent in the youth mental health state of emergency and help eliminate some of the burden on clinicians in implementing a prevention curriculum.

**Conclusions**

In conclusion, this study demonstrates the positive reception of technology—and the TIPS app in particular—in adolescent trauma-focused treatment, with clinicians, adolescents, and caregivers recognizing the benefits of engagement, accessibility, and user-friendliness of this novel mHealth tool. The qualitative themes shed light on the strengths of the app’s content, areas for improvement, esthetics, and usability as well as its potential to enhance adolescents’, caregivers’, and clinicians’ experiences during TF-CBT sessions. These findings underscore the importance of the ongoing development and refinement of digital tools in mental health care—including those that can be integrated into telehealth mental health care delivery—to better meet the evolving needs of trauma-affected adolescents and their caregivers.

**Acknowledgments**

This study was supported by K24DA039783 (principal investigator [PI]: CKD) and in part by the National Center for Advancing Translational Sciences of the National Institutes of Health under grant UL1 TR001450 (multiple PIs [MPIs]: Kathleen Brady and Patrick Flume). The manuscript preparation was also supported in part by grants from the National Institute of Mental Health (R01MH112209 [PI: CKD] and T32MH018869 [MPIs: CKD and Dean Kilpatrick]) and the National Institute on Drug Abuse (R01DA031285 and 1R01DA03288 [MPIs: CKD and Paula Riggs]; K23DA050800 [PI: AH]).

**Data Availability**

The data are available from the study principal investigator (CKD) upon request.

**Conflicts of Interest**

None declared.

**References**


18. Thielemann JF, Kasparik B, K...


36. Home page. Medical University of South Carolina. URL: https://tcfbl2.musc.edu/ [accessed 2023-12-10]


72. Claridge N, Kirakowski J. Website analysis and measurement inventory questionnaire. WAMMI. URL: http://www.wammi.com/index.html [accessed 2023-09-06]
77. Coady JA. Youth substance use prevention month. Substance Abuse and Mental Health Service Administration. 2022. URL: https://www.samhsa.gov/blog/youth-substance-use-prevention-month [accessed 2023-12-10]

Abbreviations

- **mHealth**: mobile health
- **PTSD**: posttraumatic stress disorder
- **RCT**: randomized controlled trial
- **STI**: sexually transmitted infection
- **TIPS**: Trauma-Informed Prevention for Substance Use and Risky Sexual Behavior
- **TF-CBT**: trauma-focused cognitive behavioral therapy
- **WAMMI**: Website Analysis and Measurement Inventory

©Carla Kmett Danielson, Angela Moreland, Austin Hahn, Devin Banks, Kenneth J Ruggiero. Originally published in JMIR Formative Research (https://formative.jmir.org), 18.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Virtual Exercise in Medicine: A Proof of Concept in a Healthy Population

Barbara Le Roy, PhD; Charles Martin-Krumm, Prof Dr; Charlotte Poupon, PhD; Raphaëlle Richieri, Prof Dr Med; Eric Malbos, MD, PhD; Fanny Barthélémé; Eric Guedj, Prof Dr Med; Marion Trousselard, Prof Dr Med

1 Unité neurophysiologie du stress, Institut de recherche biomédicale des armées, Brétigny-sur-Orge, France
2 Adaptation, mesure et évaluation en santé. Approches interdisciplinaires, Metz, France
3 Vulnérabilité, capabilité, rétablissement, Unité religion, culture et société, École des psychologues praticiens, Paris, France
4 École nationale supérieure de création industrielle, Paris, France
5 Département psychiatrie, Hôpital de la Conception, Institut Fresnel, Marseille, France
6 Département de médecine nucléaire, Centre européen de recherche en imagerie médicale, Institut Fresnel, Marseille, France
7 Service de santé des armées, Paris, France
8 Réseau ABC des Psychotraumas, Montpellier, France

Corresponding Author:
Barbara Le Roy, PhD
Unité neurophysiologie du stress
Institut de recherche biomédicale des armées
Place Général Valéry André
Brétigny-sur-Orge, 91220 France
Phone: 33 123456789
Email: barbara.m.le.roy@gmail.com

Abstract

Background: Science is beginning to establish the benefits of the use of virtual reality (VR) in health care. This therapeutic approach may be an appropriate complementary treatment for some mental illnesses. It could prevent high levels of morbidity and improve the physical health of patients. For many years, the literature has shown the health benefits of physical exercise. Physical exercise in a VR environment may improve the management of mild to moderate mental health conditions. In this context, we developed a virtual environment combined with an ergocycle (the augmented physical training for isolated and confined environments [APTICE] system).

Objective: This study aims to investigate the impact of physical exercise in a VR environment.

Methods: A total of 14 healthy participants (11 men and 3 women; mean age 43.28, SD 10.60 years) undertook 15 minutes of immersive physical exercise using the system. Measures included mindfulness and immersion disposition, subjective perceptions of sensory information, user experience, and VR experience (ie, psychological state, flow, and presence).

Results: First, the APTICE system appears to be a useful tool because the user experience is positive (subscales in the AttrakDiff questionnaire: pragmatic quality=0.99; hedonic quality–stimulation=1.90; hedonic quality–identification=0.67; attractiveness=1.58). Second, the system can induce a positive psychological state (negative emotion, P=.06) and an experience of flow and presence (P values ranging from <.001 to .04). Third, individual immersive and mindful disposition plays a role in the VR experience (P values ranging from <.02 to .04). Finally, our findings suggest that there is a link between the subjective perception of sensory information and the VR experience (P values ranging from <.02 to .04).

Conclusions: These results indicate that the device is well accepted with positive psychological and exteroceptive outcomes. Overall, the APTICE system could be a proof of concept to explore the benefits of virtual physical exercise in clinical medicine.

(KEYWORDS) countermeasures; mental health; physical activity; virtual reality; user experience

 JMIR Form Res 2024 | vol. 8 | e45637
doi: 10.2196/45637

https://formative.jmir.org/2024/1/e45637

(p. 323)
Introduction

The Roman poet Juneval wrote “mens sana in corpore sano” (a healthy mind in a healthy body) [1]. Still relevant today, it has never made as much sense. The body and the mind seem to be indivisible, truly part of a whole [2].

Virtual Reality

In recent years, virtual reality (VR) has been recognized as a new approach to health [3-6] that seeks to connect the body and mind [7]. The term was first used by Jaron Lanier in 1986 to refer to an advanced technological interface in which the user interacts with a 3D environment that is generated by a computer to simulate real-world experiences [8,9]. The tool can simulate reality and stimulate the body’s senses in ways that are only limited by our imagination. It creates a new space-time that is halfway between the real and the unreal, pushing back the boundaries of reality and experimenting with new paradigms that we would not otherwise have access to [10,11]. Thus, VR goes beyond a simple simulation of the external world. The modulation of interoceptive, exteroceptive, and vestibular information leads the participant to create a representation of their own body. This conceptualization is described as the body matrix, which refers to the multisensory representation of the body in the brain and the space directly around the body [12]. Through VR, it is possible to induce the illusion of being and moving in a fake body. This interstice allows individuals to perceive, interpret, and interact with their environment through an internal representation of the world [13]. Repeated VR use may stimulate changes in the brain based on neuroplasticity mechanisms [14]. Riva et al [7] noted that the effects may be heightened by immersive VR systems and the induced sense of presence in the surrounding virtual environment.

Immersion is a characteristic of VR systems and is created when the virtual environment replaces the user’s sensory stimuli with virtual sensory stimuli. Through immersion, it is possible to induce the sense of presence. Multisensory integration generates a feeling of being there and can sometimes lead to the illusion of being in an alternative body [15,16]. Slater [17] defined presence as “the strong illusion of being in a place despite the sure knowledge that you are not there.” Thus, participants have the strong illusion of being in the virtual environment and being able to perceive what is happening in it such as the virtual precipice. However, they consciously know that this is only a perceptual illusion not a reality [18]. Presence is related to flow, which refers to “the holistic sensation that people feel when they act with total involvement” [19]. It is a psychological state corresponding to enjoyment, cognitive absorption, and distortions in time perception. The literature on VR highlights the influence of immersion, induced by VR systems, on both presence and flow in the virtual environment [20,21].

Nevertheless, interindividual differences have been noted regarding both presence and flow. One relevant factor is mindfulness disposition (MD). MD is characterized by the awareness that emerges when paying purposeful attention to the present moment and responding nonjudgmentally to the unfolding experience [22,23]. It is associated with a protective function in both a healthy population and among patients [24,25]. A recent study by Lefranc et al [26] highlighted that high MD is associated with better positive emotions, interoception, and subjective extrasensory acuity. Top-down conceptual representations and bottom-up multisensory inputs contribute to body awareness. Moseley et al [12] suggested that these representations be integrated with exteroceptive data in the body matrix.

Over the years, VR has become increasingly accessible. It has been particularly beneficial in the field of medicine, whether in the context of medical training, surgery, the treatment of certain neurodegenerative diseases, rehabilitation, pain management, or cognitive and psychological disorders [11,27-34]. The literature shows the value of using VR as a therapeutic tool to treat mental disorders such as anxiety, depression, posttraumatic stress disorder, and phobias [8,31,33-38]. Antidepressants, such as selective serotonin-norepinephrine reuptake inhibitors, or benzodiazepines are the first-line treatment for anxiety symptoms in patients while cognitive behavioral therapy has been found to be effective in reducing them [39-41]. VR interventions such as exposure therapy have been shown to be effective as a coadjuvant in mental illness and appear to have the same effects as drug treatments, although the results take longer to become apparent [34]. Used as a complementary therapy, VR may have many advantages, including the ability to recreate a realistic traumatic environment under controlled conditions, which can be complex in vivo [33,42,43]. Most studies show that participants have a high degree of acceptance, and VR use is consistent with postintervention improvements in symptom awareness; a decrease in depressive symptoms; greater motivation to exercise; and better enjoyment, engagement, and affect, particularly in clinical populations [35,37]. VR therapy can stimulate emotion (notably fear), as the participant has the feeling of being present in the unreal environment [18,44]. Thus, it appears to be an innovative nondrug supplement to other treatments that can be demanding for the patient and may have side effects. Although the quality of the technology may play a role in positive outcomes [45], it appears to be an interesting new tool that poses no serious threat to participants [46].

The Potential of Immersive Physical Activity

In recent years, an increasing body of the literature has investigated the power of immersive physical activity. Physical activity preserves health and protects individuals from many pathologies [47-49]. It can be defined as “any bodily movement produced by skeletal muscles that results in the expenditure of more energy than the resting metabolism” [50]. One of the components of physical activity is physical exercise, understood as “planned, structured, repetitive physical activity whose objective is to improve or maintain one or more components of physical fitness” [50]. For many years, the literature has shown the benefits of physical activity on health, not only physical (ie, reduced mortality, reduced risk of cardiovascular pathologies, reduced incidence of cancer, or weight maintenance) but also cognitive (ie, improved cognitive function, improved sleep, or reduced risk of dementia) and psychological (ie, reduced signs of anxiety and depression or reduced risk of depression), both in the general population (ie, adults, children, and older adults) and in the context of various chronic diseases and health
conditions [47-49,51]. However, it is only recently that the scientific community has begun to take an interest in the biological and physiological mechanisms underlying these outcomes [52,53]. People with mental illness often exhibit disrupted sensory processing and perception [54]. Thus, physical activity therapy can be both a physical and psychological countermeasure. However, compliance is a key issue as regular practice is necessary for optimal mental illness management.

Few studies have examined the use of VR in this context, although the pioneering work of Plante et al [55-57] seems to indicate real benefits in terms of well-being, particularly in women [56]. The addition of VR has been found to enhance mood, increase enjoyment and energy, reduce tiredness, enhance motivation and confidence, and increase compliance [57,58]. Enjoyment may play an important role in the benefits gained from exercise [58].

In recent years, there has been an increase in the number of studies that encourage the practice of sports to prevent anxiety disorders and protect against anxiety and depression [59,60]. A recent study demonstrated its importance in the context of the COVID-19 pandemic, where it was able to improve well-being through improved physical and cognitive outcomes and limit psychological disorders related to isolation and confinement [61]. Thus, the literature suggests that VR coupled with physical activity may be a useful way to improve the symptomatology of individuals with anxiety disorders, posttraumatic stress disorder, and depression [61]. Furthermore, many studies have highlighted the ability of natural environments to induce positive emotions, promote well-being, reduce anxiety, improve self-esteem, and reduce negative emotions (ie, fatigue, confusion, tension, depression, and anger-hostility) compared with urban or indoor environments [62,63]. The same observation has been made in VR environments [64]. A virtual environment that offers physical activity in a natural setting seems to have the potential to improve the benefits of VR, especially for people with mental illness [65,66].

Gaps in the Literature and Objectives of the Study

Many of the systematic reviews and meta-analyses that have been carried out have important limitations, notably related to differences in technology. There is also a lack of longitudinal studies on the long-term effects of VR. Most studies are one-shot experiments that evaluate its benefits before and after the intervention. Evaluation itself is problematic as subjective measures (questionnaires) are typically used and few studies have measured physiological effects (ie, heart rate variability, heart rate, and electrodermal activity). As it can be complex to overcome these gaps, caution is advised in interpreting any results or conclusions [33,35,62,65,67,68]. Given these gaps in the literature, there is a need for more rigorous testing. Any evaluation should be based on three assessment criteria: (1) the activity does not duplicate other countermeasures; (2) it must improve the experience of sport and thus increase its attractiveness (especially relevant for patients with depression) [55,56]; and (3) immersion must provide a multimodal sensory input to the user [69-72]. The benefits of multisensory stimulation have been demonstrated in the context of cognitive and sensorimotor rehabilitation [73] and emotion regulation [10,74].

Thus, the aim of this preliminary proof-of-concept study was to investigate the association between VR and physical exercise in a virtual natural environment to improve the psychological state of healthy participants and the underlying processes, before evaluating its benefits in clinical medicine. We measured the user experience (UX) and evaluated 3 hypotheses:

- **Hypothesis 1**: positive changes in psychological state are associated with flow and presence during the session in the VR environment.
- **Hypothesis 2**: both MD and immersion disposition are positively related to change in the participant’s psychological state, flow, and presence.
- **Hypothesis 3**: there is a relation between the subjective evaluation of sensory information, immersive disposition, and mindful disposition and psychological change.

### Methods

#### Ethical Considerations

This study was approved by the Minarm Ethical Committee (N 125 132/MIP/DGA/MINARM). Written consent was obtained from all participants in accordance with the Declaration of Helsinki and subsequent amendments.

#### Participants

In total, 14 volunteers (3 women and 11 men), who were declared medically fit, were recruited during the 3 innovation open days at the French Armed Forces Biomedical Research Institute in 2019. They ranged in age from 22 to 59 (mean 43.28, SD 10.60) years and were either working for the French Armed Forces Biomedical Research Institute (n=9) or the French Football Federation (n=5). See Table 1 for the demographic information. The participants were recruited by email and contacted to determine whether they met the inclusion and exclusion criteria. If eligible, they were assigned an appointment for the laboratory session. All participants were asked to abstain from exercise on the day of their participation to ensure that the results were due to the experiment. The inclusion criteria were based on the following: affiliation to a health care system (social security); age between 18 and 75 years; and no history of neurological or cardiovascular disease, diabetes, or medications that could affect the response. Exclusion criteria included pregnancy, the presence of a contraindication to VR (people who had experienced anxiety or nausea during a VR experience, photosensitive epilepsy, vestibular disorder, or severe myopia >3.5 diopters).
Table 1. Sociodemographic characteristics of participants (N=14).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y), mean (SD)</td>
<td>43.28 (10.60)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>11 (78.57)</td>
</tr>
<tr>
<td>Women</td>
<td>3 (21.42)</td>
</tr>
<tr>
<td>Screen time, mean (SD)</td>
<td></td>
</tr>
<tr>
<td>Professional</td>
<td>300.00 (164.73)</td>
</tr>
<tr>
<td>Personal</td>
<td>111.42 (63.95)</td>
</tr>
<tr>
<td>Physical activity, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>10 (71.42)</td>
</tr>
<tr>
<td>No</td>
<td>4 (28.57)</td>
</tr>
<tr>
<td>Video games, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4 (28.57)</td>
</tr>
<tr>
<td>No</td>
<td>8 (57.14)</td>
</tr>
<tr>
<td>Ocular correction, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>10 (71.42)</td>
</tr>
<tr>
<td>No</td>
<td>4 (28.57)</td>
</tr>
</tbody>
</table>

Participants who practiced a physical activity or engaged in video games completed the Addictive Intensity Evaluation Questionnaire (AIEQ). The analysis found that 10 out of 14 (85%) participants engaged in physical activity (mean 31.00, SD 6.20) and 4 (35%) played video games (mean 28.20, SD 14.34). No addictive behaviors were found among the participants in either of these modalities.

Augmented Physical Training for Isolated and Confined Environments

This proof-of-concept study is based on the augmented physical training for isolated and confined environments (APTICE) system. The aim of the system is to use physical exercise in a VR environment to improve the well-being of patients with depression. It is composed of a VR - enabled cycle ergometer (VirZOOM Bike Controller) and a VR-based head-mounted display (Oculus Rift CV1, Oculus VR), which provides visual and auditory inputs. The VR application was developed by GAMIT (Petit-Quevilly) and ran on an Asus A15 TUF566IU-HN326T laptop with an AMD Ryzen 5 4600H 16 GB processor, a 512 GB solid state drive, and an Nvidia GeForce GTX1660 Ti 6 GB graphics card. The VR environment consisted of natural areas of forests and mountain plains (Figure 1). See Multimedia Appendix 1 for further images of the APTICE device.

Figure 1. Natural virtual environment images. (A) Forest with stretches of water. (B) Mountain plain with sheep.
Data Collection

Population Overview
A 7-item sociobiographical questionnaire was developed to collect standard sociodemographic data such as gender, age, hobbies, physical activity, video game use, and VR experience. The AIEQ evaluated addictive practices [75]. Two versions were used: the 14-item AIEQ-g that measures the intensity of video game playing and the risk of its problematic use and the 14-item AIEQ-s that measures sports practice and the risk of its problematic use.

UX of the APTICE Device
The UX of the APTICE device was assessed using the 10-item AttrakDiff questionnaire, which evaluates the hedonic and pragmatic qualities of interactive systems [76]. It is divided into 4 subscales: pragmatic quality, hedonic quality–identity, hedonic quality–stimulation, and attractiveness. Values close to the mean (from 0 to 1) are considered standard values. They indicate that the device meets its objectives with no negative impacts on the user.

Psychological Questionnaires
Two questionnaires were used to evaluate psychological dispositions. The 14-item Freiburg Mindfulness Inventory was used to measure MD [77]. It is divided into 2 subscales: acceptance and presence. Immersion disposition was assessed using the 18-item immersive tendencies questionnaire, which is divided into 4 subscales: focus, involvement, emotions, and games [78]. Two questionnaires were used to evaluate psychological state. First, the 12-item Scale of Positive and Negative Experience (SPANE) questionnaire assessed subjective feelings of well-being [79]. The overall scale is divided into 2 subscales: positive and negative emotions. Second, the 20-item Activation-Deactivation Adjective Checklist (AD-ACL) evaluates the level of awareness and emotional disposition [80]. This is divided into 2 dimensions: energetic arousal (from energy to tiredness) and tense arousal (from tension to calmness). The energetic arousal is further divided into 2 subscales—general activation and deactivation—while the tense arousal is subdivided into general tenseness and calmness.

Subjective Evaluation of the Quality of Sensory Information
We developed the Personal Evaluation of Six Senses questionnaire to assess subjective perceptions of vision, sound, touch, olfaction, taste, and equilibrium. Participants evaluated the accuracy of their perceptions from each of their 6 senses using a ranked scale running from 1 to 6.

The VR Experience
The VR experience was assessed using the 12-item Educational Flow Questionnaire (EduFlow2), which measures flow [81]. It is divided into 4 dimensions: cognitive control, immersion and time transformation, loss of self-consciousness, and autotelic experience. Cognitive absorption (a summary of the first 3 dimensions) was added as the fourth scale. The 24-item Presence Questionnaire assessed presence [82]. It is divided into 7 subscales: realism, possibility of action, quality of interface, possibility of examination, self-evaluation of performance, sounds, and haptic. APTICE device sickness was assessed using the 16-item Simulator Sickness Questionnaire [83]. It is divided into 2 subscales: nausea and oculomotor.

Procedure
The experimental protocol is illustrated in Figure 2. Upon arrival, the participants were asked a few questions to ensure they met the inclusion criteria and signed the consent form. They then completed a series of questionnaires in the following order: sociobiographical questionnaire, AIEQ-g, AIEQ-s, Freiburg Mindfulness Inventory, Personal Evaluation of Six Senses, 18-item immersive tendencies questionnaire, SPANE, and AD-ACL. Then, they engaged in a moderate-intensity bout of exercise in a natural environment for 15 minutes while wearing the VR headset. They could choose their trajectory along various predefined paths and, by turning their head, obtain a 360° view of the virtual environment. At the end of the session, they were asked to complete another series of questionnaires in the following order: SPANE, AD-ACL, EduFlow2, 24-item Presence Questionnaire, AttrakDiff, and Simulator Sickness Questionnaire.
Figure 2. The augmented physical training for isolated and confined environments (APTICE) experimental protocol. AD-ACL: Activation-Deactivation Adjective Checklist; FMI: Freiburg Mindfulness Inventory; ITQ-f: Immersive Tendencies Questionnaire; SPANE: Scale of Positive and Negative Experience; SSQ: Simulator Sickness Questionnaire; VR: Virtual Reality.

**Statistical Analysis**

Statistical analyses were performed using the RStudio software (version 1.2.5001). Descriptive statistics are expressed as mean (SD). The Shapiro-Wilk test was used to determine whether the data were normally distributed. The effects of the APTICE device experience on emotional and activation-deactivation states were assessed as follows: a t test (2-tailed) for pre-post comparisons and parametric data or the Mann-Whitney U test for nonparametric data. Kendall correlations were run to explore the relationship among virtual exercise, subjective sensory accuracy, and VR experience. For all analyses, significance was set at $P<.05$. Trends were considered when $.05<P<.10$. Deltas were calculated to compare the temporal impact of the experience measured using the SPANE questionnaire and the AD-ACL.

**Results**

**The UX**

The APTICE tool was assessed in terms of UX. Participants reported a positive experience measured as pragmatic quality, hedonic quality--stimulation, hedonic quality--identification, and attractiveness (Figure 3). The scores were particularly high for hedonic quality--stimulation and attractiveness. No participant reported any cybersickness.
Figure 3. AttrakDiff’s subscales. Values close to the mean (from 0 to 1) are considered standard and indicate that the device meets its objectives with no negative impacts on the user. However, they also suggest that improvements could be made to the system to obtain high positive values. Values outside this neutral zone are considered positive (1-3) or negative (−1 to −3). X1: pragmatic quality; X2: hedonic quality (stimulation); X3: hedonic quality (identification); X4: attractiveness.

Relationships Between Psychological Assessments, Exteroception, and VR Experience

Overview
Multimedia Appendix 2 summarizes the significant correlations between the tested variables.

Hypothesis 1: The VR Experience (Change in Psychological State, Flow, and Presence)
The analysis of emotional and arousal states only identified a trend for negative emotions. Participants tended to have fewer negative emotions after the APTICE experiment ($t_{12}=2.06$, $P=.06$).

There were significant positive and negative correlations between flow and presence. Participants who scored high for possibility to examine also scored high for flow cognitive control ($\tau=0.45$, $P<.001$), flow cognitive absorption ($\tau=0.67$, $P=.001$), and flow immersion and time transformation ($\tau=0.55$, $P=.01$). Participants who scored high for possibility to act also scored high for flow cognitive absorption ($\tau=0.58$, $P=.004$), flow cognitive control ($\tau=0.76$, $P<.001$), flow immersion and time transformation ($\tau=0.58$, $P=.006$), and flow-autotelic experience ($\tau=0.58$, $P=.001$). As realism increased, flow cognitive control also increased ($\tau=0.52$, $P=.01$). However, as haptic increased, flow loss of self-consciousness ($\tau=-0.52$, $P=.02$) decreased.

Concerning change in psychological states related to flow and presence, our results suggest that there is no correlation between change in emotional state (measured with the SPANE questionnaire) and either flow or presence. However, there were significant negative correlations between flow and changes in activation-deactivation states (measured using the AD-ACL). An increase in tense activation (positive delta) was associated with lower scores for flow immersion and time transformation ($\tau=-0.46$, $P=.04$) and flow-autotelic experience ($\tau=-0.52$, $P=.01$). No correlation was found between presence and flow, and there were no changes in activation-deactivation.

Hypothesis 2: Disposition and the VR Experience (Change in Psychological State, Flow, and Presence)
No relationship was observed between immersive disposition and MD for any subscale.

The analysis found a significant positive correlation between MD and presence. More precisely, higher MD-acceptation was associated with a higher score for possibility to examine ($\tau=0.49$, $P=.02$). There was also a significant positive correlation between MD and flow. High scores for MD-acceptation were associated with high scores for flow cognitive control ($\tau=0.45$, $P=.03$).

Finally, there was a significant positive correlation between immersion and flow. Specifically, high scores for flow loss of self-consciousness were slightly associated with high scores for involvement ($\tau=0.54$, $P=.01$).

Concerning disposition and the VR experience, the analysis found no correlation between change in emotional state and either immersive or mindful disposition. Significant positive and negative correlations were found between immersion and change in activation-deactivation. An increase in tense activation (positive delta) was associated with higher scores for games ($\tau=0.45$, $P=.04$). However, an increase in general activation
(positive delta) was associated with lower scores for *involvement* ($\tau = -0.52$, $P = .02$).

**Hypothesis 3: Subjective Exteroceptive Accuracy, Disposition, and the VR Experience**

The analysis found no relation between the subjective exteroceptive evaluation and changes in emotional and activation states, presence, or MD. However, significant positive and negative correlations were observed between immersion and subjective acuity.

Correlation matrices for immersion and subjective acuity variables are shown in **Figure 4**.

**Figure 4.** Correlation matrices for immersion and subjective acuity variables. Distributions are shown on the diagonal. Trend curves are shown at the bottom of the diagonal scatter plots. The top diagonal shows correlation coefficients and significance levels. A1: olfaction; A3: vision; I1: focus; I2: involvement. *>.99, **.10, ***.05, ****.01.
Increased involvement was associated with higher subjective visual acuity ($\tau=0.48, P=0.03$). In contrast, an increase in focus was associated with lower subjective smell acuity ($\tau=-0.43, P=0.04$). Low scores for subjective hearing were associated with high scores for flow cognitive absorption ($\tau=-0.43, P=0.04$) and flow immersion and time transformation ($\tau=-0.47, P=0.03$). Similarly, low scores for subjective taste were associated with high scores for flow cognitive control ($\tau=-0.49, P=0.02$).

Discussion

Principal Findings

The main aim of this proof-of-concept study was to investigate the effect of VR associated with physical activity on the psychological state of healthy participants before further evaluation of patients with depression in a randomized controlled trial. This exploratory study evaluated a new device, named APTICE, which couples physical exercise with a VR headset. This pilot feasibility study proposed variables of interest, which will form the basis for our next randomized controlled trial. The latter will investigate clinical and neurofunctional subtraits in a population affected by depression using VR associated with physical activity. Results from this study provide new insights into the benefits of this type of technology when used in clinical medicine to improve health.

A Positive UX Experience

As Hassenzahl et al [84] demonstrated, the evaluation of the hedonic and pragmatic qualities of a system is known to influence overall perceptions of its attractiveness. Understanding the UX is crucial in the design of a new device, which is often ignored. The participants in our study were very positive regarding both the hedonic quality–stimulation and the attractiveness of the device. However, pragmatic quality and hedonic quality–identification scores were lower. Furthermore, the responses were the most disparate for these 2 dimensions. Hedonic quality–stimulation was associated with ideas such as outstanding, impressive, exciting, or interesting. Although the response to the UX appears to be positive, there is room for improvement. The relatively low scores for hedonic quality–identification are not surprising, as this aspect relates to the ability of the system to reflect the user’s identity. Similarly, pragmatic quality needs to be improved with a focus on usability. Both the appropriateness of the functionality and its accessibility need further attention. However, this short 15-minute experiment allowed us to conclude that the APTICE system meets its development and quality objectives—specifically, to design a device that supports physical exercise in VR. In the longer term, we will need to consider how to improve it, particularly in light of the technological development that has taken place since its creation.

Psychological Changes Induced by the APTICE Device

Our main hypothesis was that physical exercise in a VR environment could create a positive experience, measured as psychological and sensory feedback from the participants. Consistent with the literature, our initial results suggest that the APTICE device experience decreases negative emotions [44,85–89]. However, our first hypothesis (that the APTICE device would induce a positive psychological state and an experience of flow and presence) was only partially confirmed. The literature [7,17] notes that presence and flow are usually positively linked, although a negative correlation has been found between haptic presence and loss of self-consciousness in flow experiments. In the absence of a meaningful haptic system, interactions with objects in the VR environment can widen the gap between actual and virtual realities [90]. In our experiment, haptic feedback from the interaction with the ergocycle did not reflect reality, which suggests its key role in inducing flow. For example, there was no body movement when going around the bends and almost no return on effort. The poor quality of the correspondence between the virtual exercise environment and reality could explain the absence of a change in positive emotions.

Our initial results suggest a close relationship between the quality of the technology and the VR experience. This is all the more important as flow (characterized by a deep involvement and absorption in an activity) promotes a state of inner well-being and positive emotions [91,92]. Overall, our results suggest that practicing a physical activity in a VR setting could be used to improve psychological outcomes. According to previous studies [60,61,66], the APTICE device may have potential benefits for patients, especially those with mental illness. The literature also shows that natural scenes support a positive psychological state both in general [65,86] and in the treatment of mental illness [67,93]. This is in line with the reduction in negative emotions in individuals following our study’s APTICE session. Although APTICE needs improvement, both the positive response to the UX and its effect on the user’s psychological state suggest that regular use may have a positive impact on mental health.

Relationships Between Disposition and VR Experience

Our results partially confirm our second hypothesis, which focused on the impact of immersive disposition and MD on the VR experience. We found no relationship between immersive disposition and MD in our sample. Immersive disposition is used to evaluate the potential to immerse a subject in a situation, whereas MD is characterized by the ability to be in the here and now. Therefore, it is possible that these 2 dimensions are unrelated. Our experiment showed that the involvement subscale of immersive disposition was associated with a loss of self-consciousness in terms of flow effect. An individual’s interest in a target object [94] or their motivational state in relation to a target object [95] has been described as a condition for flow experience in VR [96]. Furthermore, our experiment found that immersion was unrelated to presence, which conflicts with the literature [15,37]. A key difference compared with earlier work is that our participants were asked to make a physical effort. It is possible that this effort counteracted their immersive disposition. If we turn to the relationship among MD, presence, and flow, acceptance seems to be the most relevant dimension. Acceptance consists of accepting inner events such as emotions, thoughts, or beliefs as they are felt [97]. It does not mean resignation but rather perceiving one’s own experience with an attitude that acknowledges it, rather than judging it as either good or bad. Thus, the ability to accept what is happening now may be a more useful way to examine

https://formative.jmir.org/2024/1/e45637

JMIR Form Res 2024 | vol. 8 | e45637 | p.331
Le Roy et al
presence and cognitive control than simply being in the here and now. Collectively, these results suggest that physical exercise in VR may be improved by acceptance, which enhances the feeling of presence.

**APTICE Device and Exteroceptive Modulations**

Our final hypothesis concerning the relationship between subjective exteroceptive perceptions of sensory information and physical exercise in the VR experience was exploratory. On the one hand, our results show that there is an assumption that information provided by all 5 senses may help the user to become immersed in the experience of where they are, whom they are with, and what they are doing. The feeling of a real experience gives rise to presence. On the other hand, mindful participants pay more attention to information from their bodies, leading to better adaptation to the environment [98]. Using functional magnetic resonance imaging, Farb et al [99] identified several brain regions associated with mindfulness. In particular, they found that deactivation of the medial prefrontal cortex and increased activation of parietal areas were associated with proprioception and sensory–motor body experiences. Mehling [100] reported the use of external stimulation when attempting to understand how felt sensations are used internally to regulate stress or attention. Such information is integrated and linked to the person’s emotional state as a function of whether the body is experienced as safe [98,100].

Our results suggest that subjective preferences in exteroception-perception are linked to the experience of physical exercise in the VR environment. Furthermore, they show that immersion is correlated with subjective visual acuity. The participants in our experiment were cycled in a virtual environment based on natural visual information. Unsurprisingly, high scores for subjective visual acuity were associated with flow. Many studies have highlighted the potential of external sensory information to enrich the lived experience [101-104]. Exteroceptive information can generate intense emotional processes [105] and flavor manipulation within VR [101-104]. However, the evidence is weak, and it is also possible that such an environment may inhibit VR experiences because of its limited capacity to provide wider sensory inputs [105]. Another outcome of our study was that individual preferences may play a role in the VR experience. Our findings showed that this experience is negatively associated with all forms of external sensory stimulation (ie, hearing, taste, and smell) except vision. This suggests that other senses are partially inhibited, and only vision is recruited on a large scale. Vision is an essential component of the APTICE experience.

In this context, Slater and Usoh [106] suggested that an individual’s experience is encoded by visual, auditory, and kinesthetic systems of representation. Depending on the context, the person will naturally tend to favor one system over another. However, the latter authors noted that the visual system predominates in individuals who report a higher sense of presence and those who process information in the first person. Thus, individual characteristics may be a key factor in any experiment. Overall, our study suggests that the APTICE system may alter multisensory representations during physical exercise.

Future studies should address this issue, which remains unexplored.

**Future Clinical Applications**

VR technologies appear to complement established approaches to mental health care. Its association with physical activity makes it an interesting new approach that merits further investigation. Furthermore, the use of VR in health care is expanding rapidly. There are many new opportunities in clinical medicine, including mental illness, where VR may be an alternative treatment [4,5,107,108]. Our findings validated the impact of physical exercise in a VR environment on negative emotions in a healthy population. Although our results should be interpreted with caution, because of the small sample size, they highlight the importance of better understanding the processes involved in healthy participants. Beyond the efficacy of interventions to determine which populations might benefit from VR combined with physical activity, it is important to understand the processes that predispose this state in healthy individuals. Further studies with larger sample sizes are required to evaluate the role of these processes in clinical research. Thus, the next step is to study clinical and neurofunctional subtracts in a population with depression before proposing the tool as a countermeasure (ID-RCB: 2020-A03415-34) for this population and other people in health care. There is an untapped opportunity to use VR as a prevention tool and to target the processes that make an individual poorly adapted to the environment. This is particularly the case for people who work in challenging confined and isolated environments or extreme and unusual environments [65].

**Limitations**

This study has 4 main limitations. The first and most important factor is the small sample size. This study was intended to be a pilot feasibility study that will support a future controlled randomized trial. In this context, it validated the usefulness of the APTICE system and highlighted the interaction between the variables of interest. In the next phase of our work, we will launch a larger clinical study of participants with depression. The second limitation relates to the use of subjective self-report measures. An objective sensory evaluation needs to be developed for healthy participants, which would help researchers to better investigate the human-body relationship. Subjective variables should be combined with physiological measures, such as heart rate variability. Third, our results cannot be generalized because the study population was recruited from among armed forces personnel and footballers, who are usually different from the general population in terms of fitness and psychological state. Finally, the last limitation concerns the VR equipment used in our experiment, which is becoming dated. A new version of the Oculus headset is already available, with a better graphics interface.

**Conclusions**

This exploratory proof-of-concept study investigated some of the processes implicated in physical exercise in a VR environment with the aim of better understanding their relationship with psychological state in a sample of healthy individuals. It represents the first step in a larger randomized...
controlled trial that will investigate clinical and neurofunctional
subtracts in a population with depression. Our results suggest
that the APTICE environment can change negative emotional
states, consistent with the experiences of flow and presence.
Moreover, our findings demonstrate that immersive and mindful
disposition play an important role in the VR experience. Finally,
they also suggest that the subjective exteroceptive perception
of sensory information may be a key aspect and seems to
indicate that one sense may prevail over another at the level of
the individual. Our study has several implications for clinical
medicine: (1) VR can help enhance and reinforce the beneficial
actions of physical activity; (2) APTICE is a promising system
and may be effective in improving mental health; and (3)
APTICE has the potential to be used as an alternative treatment
to drugs and to improve quality of life. However, many
questions remain unanswered, and further work is needed to
exploit the potential of VR associated with physical activity
both as prevention and treatment.

Acknowledgments
The authors wish to thank Eric Malbos. The authors thank the French Armed Forces Biomedical Research Institute and French
Football Federation for their involvement. They would like to thank Florent Samat and Nicolas Greverie from GAMIT France
for the development of the APTICE system. Finally, the authors would like to thank the Centre Nationales des Études Spatiales
and the Direction Générale de l’Armement for their support.

This research was funded by the Direction Générale de l’Armement (N 2016/017/S).

Data Availability
The data sets generated and analyzed during this study are not publicly available because they are the property of the French
Armed Forces Health Service. Although data are not available to the public, they are available from author MT upon reasonable
request.

Authors’ Contributions
BLR, CMK, CP, RR, EM, FB, EG, and MT conceptualized the study and developed the methodology. All authors actively
contributed to the experimental process. BLR, CMK, and MT wrote the original manuscript. BLR conducted the formal analysis.
All authors have read and approved the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Augmented Physical Training for Isolated and Confined Environments (APTICE) system description.
[DOCX File, 2387 KB - formative_v8i1e45637_app1.docx ]

Multimedia Appendix 2
Significant correlations between tested variables as a function of the 3 hypotheses.
[DOCX File, 31 KB - formative_v8i1e45637_app2.docx ]

References
Res 2017;9(9):3867-3880 [FREE Full text] [Medline: 28979666]
33401596]


https://formative.jmir.org/2024/1/e45637

JMIIR FORMATIVE RESEARCH
Le Roy et al
Québécoise pour la Recherche en Psychologie (SQRP); November 1-3, 2002; Trois-Rivières, Canada. [doi: 10.1037/h0081306]


Abbreviations

AD-ACL: Activation-Deactivation Adjective Checklist
AIEQ: Addictive Intensity Evaluation Questionnaire
APTICE: augmented physical training for isolated and confined environments
MD: mindfulness disposition
SPANE: Scale of Positive and Negative Experience
UX: user experience
VR: virtual reality

Edited by T Leung; submitted 10.01.23; peer-reviewed by R Menhas, Y Theodorakis; comments to author 28.03.23; revised version received 19.05.23; accepted 31.08.23; published 22.01.24.

Please cite as:
Virtual Exercise in Medicine: A Proof of Concept in a Healthy Population
JMIR Form Res 2024;8:e45637
URL: https://formative.jmir.org/2024/1/e45637
doi:10.2196/45637
PMID:38252484

©Barbara Le Roy, Charles Martin-Krumm, Charlotte Poupon, Raphaëlle Richieri, Eric Malbos, Fanny Barthélémy, Eric Guedj, Marion Trousselard. Originally published in JMIR Formative Research (https://formative.jmir.org), 22.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Experiences Among Patients With Cystic Fibrosis in the MucoExocet Study of Using Connected Devices for the Management of Pulmonary Exacerbations: Grounded Theory Qualitative Research

Maxime Morsa, PhD; Amélie Perrin; Valérie David, Dr med; Gilles Rault, Dr med; Enora Le Roux, PhD; Corinne Alberti, PhD; Rémi Gagnayre, Dr med; Dominique Pougheon Bertrand, PhD

1Adaptation, Resilience and Change Research Unit, University of Liège, Liège, Belgium
2Laboratory Health Promotion and Education (UR3412), Sorbonne Paris North University, Bobigny, France
3Paediatrics CF Centre, Nantes University Hospital, Nantes, France
4Laboratoire Educations et Promotion de la santé, Université Sorbonne Paris Nord, Bobigny, France
5Institut National de la Santé et de la Recherche Médicale (UMR 1123 ECEVE), Université de Paris, Paris, France
6Unit of Clinical Epidemiology (CIC-EC 1426), Hôpital Universitaire Robert Debré, Assistance Publique des Hôpitaux de Paris, Institut National de la Santé et de la Recherche Médicale, Paris, France

Corresponding Author:
Maxime Morsa, PhD
Adaptation, Resilience and Change Research Unit
University of Liège
1 Place des Orateurs
Liège, 4000
Belgium
Phone: 32 4 3662272
Email: maxime.morsa@uliege.be

Abstract

Background: Early detection of pulmonary exacerbations (PEx) in patients with cystic fibrosis is important to quickly trigger treatment and reduce respiratory damage. An intervention was designed in the frame of the MucoExocet research study providing patients with cystic fibrosis with connected devices and educating them to detect and react to their early signs of PEx.

Objective: This study aims to identify the contributions and conditions of home monitoring in relation to their care teams from the users’ point of view to detect PEx early and treat it. This study focused on the patients’ experiences as the first and main users of home monitoring.

Methods: A qualitative study was conducted to explore patients’ and professionals’ experiences with the intervention. We interviewed patients who completed the 2-year study using semistructured guides and conducted focus groups with the care teams. All the interviews were recorded and transcribed verbatim. Their educational material was collected. A grounded analysis was conducted by 2 researchers.

Results: A total of 20 patients completed the study. Three main categories emerged from the patients’ verbatim transcripts and were also found in those of the professionals: (1) task technology fit, reflecting reliability, ease of use, accuracy of data, and support of the technology; (2) patient empowerment through technology, grouping patients’ learnings, validation of their perception of exacerbation, assessment of treatment efficacy, awareness of healthy behaviors, and ability to react to PEx signs in relation to their care team; (3) use, reflecting a continuous or intermittent use, the perceived usefulness balanced with cumbersome measurements, routinization and personalization of the measurement process, and the way data are shared with the care team. Furthermore, 3 relationships were highlighted between the categories that reflect the necessary conditions for patient empowerment through the use of technology.

Conclusions: We discuss a theorization of the process of patient empowerment through the use of connected devices and call for further research to verify or amend it in the context of other technologies, illnesses, and care organizations.

Trial Registration: ClinicalTrials.gov NCT03304028; https://clinicaltrials.gov/ct2/show/results/NCT03304028
Introduction

Background

Pulmonary exacerbations (PEx) are the main cause of decline in lung function in patients living with cystic fibrosis (CF), representing the leading cause of death. Recommendations emphasize the importance of diagnosing PEx early to treat patients effectively and for them to have the best chance of regaining their previous baseline lung function after treatment [1]. Identifying warning signs of PEx requires many skills from patients daily—studies have shown that they must be able to monitor a combination of physiological parameters and patient-reported perceptions, such as weight loss, decreased spirometry, increased coughing, or increased sputum production reported daily, to diagnose PEx episodes and put in place the appropriate treatment [2,3]. Nevertheless, patients living with CF do not systematically monitor these warning signs, as few are equipped with devices to monitor variations in their physiological parameters or their perceptions over time, with the exception of patients who have received a lung transplant, who may be equipped with spirometers to detect a decrease in their respiratory function, which is a warning sign of acute rejection. However, patients need to access accurate and reliable measurements to monitor their lung function.

In recent years, a contemporary trend has emerged in health improvement and disease prevention: the “quantified self.” It refers to the quantitative measurement of various parameters linked to the state of one’s health (eg, heart rate and weight) or to lifestyle (eg, diet and physical activity) to monitor a disease or improve well-being. The premise is that one cannot improve what they cannot quantify. This quantification, which was still challenging a few years ago, has become more accessible through the development of new technologies and connected devices. These devices are connected to the internet and can collect, store, process, and transmit health-related data through sensors [4].

Connected devices can help patients gain a better understanding of disease and treatment and increase their levels of satisfaction and adherence to treatment when combined with patient education interventions [5,6]. Patient education is an empowering approach for patients with chronic diseases aiming to improve their understanding and adherence to treatment by transferring knowledge from health care providers to patients through educational workshops and also by using patients’ experiential knowledge, which helps them adjust their management of the disease in their daily lives [7]. Patient education is known to have a significant positive impact on bioclinical indicators and on the well-being of patients [8]. Connected devices would act as a learning aid for patients by promoting real-life behavioral experimentation thanks to quick (or even immediate) access to objective data and to the development of knowledge about oneself anchored in one’s memory [9]. The use of connected devices by patients in their daily lives allows them to transfer what they learned during the workshops provided by health care providers to real-life situations, thus expanding on patient education. Experiential and continuous learning is facilitated when it is supported by health care providers to learn to interpret real-life data and compare them with the data collected at the hospital.

This way, connected devices could promote the process of empowerment, a concept that is understood as the development of patients’ ability to identify and meet their own needs, solve their own problems, mobilize the necessary resources to take action, and feel that they are in control of their health and their own lives [10]. According to Funnell and Anderson [11], empowerment is a process that is facilitated by counseling, educational, or psychological techniques to help the individual take control of the day-to-day management of their illness.

Currently, data are scarce on how connected devices are used in real-life situations by patients with chronic diseases and on how they influence knowledge of oneself and of one’s body, health, and disease [12]. However, we know that the dropout rate of connected devices can be high because of how cumbersome their use may be or the fact that they are too pressing a reminder of the person’s disease in their daily life [13,14], whereas adherence is mainly observed in young people and high-income socioprofessional categories who are more familiar with new technologies [15]. People’s experiences of using such connected devices vary depending on the person, the context, and their care environment. Therefore, the assessment of health technology is now moving toward a contextualized, patient-based evidence approach. According to this approach, the evaluation of eHealth devices is based on knowledge that originates directly from patients about their experiences of health, quality of life, and health services [16]. This approach is represented internationally by the work of the Warwick Patient Experiences Framework or the National Institute for Health and Care Excellence Patient Experience Guideline Development Group [17].

Drawing from the humanities and social sciences, it is now recommended for qualitative studies to be centered on patients’ feedback to understand the processes through which connected devices facilitate their acquisition of knowledge (of the body, risks, and diseases), in particular through the intimate and empirical experiences of the quantified body translated into data [18].

Objectives

Therefore, we conducted a qualitative study with patients living with CF and with specialized CF centers in metropolitan France to explore the processes through which connected devices become an essential part of patients’ knowledge to allow them to self-manage their health and to contribute to a theory of individual patient empowerment through technology. The aim

KEYWORDS
cystic fibrosis; mobile health; mHealth; patient education; chronic disease; empowerment; devices; patients; detection; treatment; respiratory; education; monitoring; care

https://formative.jmir.org/2024/1/e38064

JMIR Form Res 2024 | vol. 8 | e38064 | p.340
(page number not for citation purposes)
of the study was to understand how patients and health care providers lived and perceived this new intervention based on connected devices associated with patient education workshops to identify the contributions and conditions of home monitoring. The work is focused on stakeholders’ experience with the intervention. This study is part of an interventional project based on the hypothesis that an intervention that combines the provision of connected devices set up with personalized alert thresholds and a patient education intervention by health care providers can enable patients with CF to detect early signs of PEx and begin managing it themselves in a timely manner. For this self-management process to lead to the implementation of appropriate patient behavior, it is assumed that the educational intervention teaches patients to identify and respond appropriately to alerts.

**Methods**

**Overview**

The MucoExocet (from the French for “Cystic Fibrosis Exacerbation Connected Devices Therapeutic Education”) study, a pilot interventional study, was conducted from 2018 to 2021 and involved 22 adults and 14 adolescents (aged >12 years) with CF to assess whether the use of connected devices was feasible and useful to detect and treat PEx early (trial registration: ClinicalTrials.gov NCT03304028). As part of the overall research project, this qualitative study explored the users’ experiences at the end of the intervention. The intervention and protocol have been extensively described previously [19]. We used the EQUATOR (Enhancing the Quality and Transparency of Health Research) standards for reporting qualitative research elaborated by O’Brien et al [20] to present our study design.

**Summary of the Intervention in Its Context**

Since 2005, a national organization associating health care providers from CF centers and patients and parents in France has been working to define the patient and parent competency framework (in pediatrics) and the associated set of educational tools. A therapeutic education tool named “Reagir en cas d’exacerbation” ("Réagir en cas d’ exacerbation") was used to support patients' and parents’ self-management of PEx episodes at home (Multimedia Appendix 1).

The intervention designed for the MucoExocet study combined the provision of connected devices with an educational program based on the React to PEx tool. It was renamed “React with CDs” and incorporates measurements from connected devices and personalized alerts (Figure 1). The goal of the intervention was to develop the patients’ (or parents’) ability to take action at the first signs of exacerbation identified through measurement deviations by connected devices. For this study, connected devices were used to collect 13 parameters, including 6 physiological parameters measured by the devices (forced expiratory volume in 1 second [FEV1], cardiac frequency, arterial hemoglobin oxygen saturation, weight, sleep duration [min/night], and physical activity [step count/d]) and 7 patient-reported perceptions described using emoticons in a journal provided by the spirometer application (trouble breathing, need for more airway clearance, increased symptoms at night, difficulty performing usual activities, greater fatigue, loss of appetite, and change in sputum [color or quantity]). At the request of both physicians and patients, the option chosen in the study was to not send the data collected via connected devices to the physician but only to the patient. However, the data could be shared during a consultation at the center or during a phone call or email exchange if the patient (or parent) wished to do so.

**Figure 1.** Design of the study intervention. CD: connected device; EW: educational workshop; M0: month 0; M3: month 3; M6: month 6; M9: month 9; M12: month 12; M15: month 15; M18: month 18; M21: month 21; PEx: pulmonary exacerbation.
Thresholds of irregular variation for patients’ parameters were calculated using the cumulative sum control chart method based on the data collected during phase 1 of the implementation of connected devices (3 months); this allowed physicians to set alert thresholds for each parameter and each patient during the first educational workshop with patients. Thus, alerts could be sent by email to the patients or parents throughout the period of routine monitoring using connected devices (12 months). An action plan was agreed upon by the physician and the patient or parent during this educational workshop to respond to alerts.

**Development of the Educational Tool and Educational Intervention in the Centers**

One of the centers played a leading role in the implementation of the study because of its leading role in therapeutic education for patients with CF in France (the CF center in Nantes). During phase 2, the physician and therapeutic patient education nurse developed the different educational tools for the 3 stages of the program (educational workshops 1-3) based on the statistical analyses of patient data and with the participation of an adult patient expert and tested these tools with a parent and an adolescent patient from their center. The tools and educational program are described in the publication cited previously [19].

Beyond the patient recruitment process, a physician and a health care provider (nurse or physiotherapist) from each investigating center were involved in handing the connected devices to the patients or parents, setting them up with the patients or parents and explaining how to use them, solving technical problems with the patients or parents with the help of the device suppliers, and participating in the interpretation of statistical analysis of the data collected during the implementation phase (3 months) to define personalized alert thresholds for their patients. They were trained in the use of the React with CDs educational tool to agree with their patients on an action plan in the event of PEx; they educated their patients in the 3 sessions provided as part of the educational program (“Your impressions of using connected devices during the implementation phase” and “Your action plan for responding to exacerbation warning signs, and Review of your action plan after 6 months of routine monitoring”). At the end of the study, health care providers participated in a focus group to report on their experience with the study and this monitoring method.

**Study Population and Study Centers**

The centers were selected by the research group on a voluntary basis among centers who had participated in the quality improvement program. They were familiar with the educational tool “React to an exacerbation.” The 7 CF centers were selected to include patients and their families of various conditions of life, economic statuses, and geographic areas (either urban or rural). Finally, the 7 centers were located in 4 different geographical areas: 3 (43%) were pediatric centers (4 patients per center), and 4 (57%) were adult centers (6 patients per center). None had previous experience with connected devices for their patients at the initiation of the study. In total, 36 patients, adults or adolescents, were included in the MucoExocet study. The sample was defined according to the recruitment capacities of the centers and the possibility of observing a saturation phenomenon in the qualitative study [21].

The recruitment process conducted by physicians in the centers was carried out based on patients’ voluntary participation and their interest in using new technologies. The participant inclusion criteria were as follows: age of ≥12 years, clinically stable condition (no PEx requiring intravenous antibiotics within the previous 4 weeks), at least one PEx within the previous 12 months, current follow-up at a participating CF center (and no plans to change centers during the course of the study), no history of having undergone solid organ transplants, prescription of at least one pulmonary medication (eg, inhaled mucolytic, inhaled or oral antibiotic therapy, or hypertonic saline), ability to speak French, ability to connect a tablet to Wi-Fi, and provision of written informed consent.

The number of individuals to be recruited took into account an estimated dropout rate of 20%. A form was offered to the patients leaving the study to identify the main reasons for their withdrawal (Multimedia Appendix 2).

**Data Collected by the Patients Using the Connected Devices**

The 13 parameters were collected using 5 different connected devices, and a few of them collected more than one parameter. During the first 3 months, data were to be collected twice a week. During the routine phase, the frequency was agreed upon between the patient and the care team from twice a week to once every 2 weeks depending on the patient’s health outcomes and life conditions. The procedure used to collect the data was explained in a document given to the patients at inclusion (Textbox 1).

Owing to the great variety of measurements taken, the time spent on the measurements was not recorded.

**Textbox 1. Procedure to collect the data.**

- Data collected without any participation from the patient (sleep, steps, and cardiac frequency): the duration of sleep was measured by the sensor under the mattress, and the step count and cardiac frequency were measured by the watch.
- Clinical data (spirometry and oxygen) required patient participation; at the end of the spirometry measurement, emoticons were presented for each of the 7 perceptions.
- Weight was expected to be measured in the morning (naked) the same day as the clinical data.

**Qualitative Data Collection**

Patients’ experiences were collected through semistructured interviews using an interview guide with 8 open-ended questions (Textbox 2), derived and adapted from validated protocols for patient narrative elicitation in outpatient care experiences [22]. The experience and workload of the care teams were explored

https://formative.jmir.org/2024/1/e38064
in focus groups using an interview guide with 5 open-ended questions (Textbox 3).

Three sources of data were collected: (1) data collected during patient or parent interviews using an interview guide with open-ended questions (Textbox 2), (2) data regarding the educational program with the physician (the documents completed by the patient and the clinician, including the personalized action plan in case of PEx), and (3) data from the focus groups with care teams at the end of the study using a semistructured guide (Textbox 3).

Textbox 2. Guide for the semistructured interviews with patients or parents.

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. For you, what are the most important aspects in the management of your respiratory exacerbations in your daily life?</td>
<td></td>
</tr>
<tr>
<td>2. How do you rate the conditions for managing exacerbations during the study (based on what is most important to you)?</td>
<td></td>
</tr>
<tr>
<td>3. Can you tell us about a positive experience you had during this study concerning the management of your exacerbations? What happened and how did it make you feel? Did you do anything in particular after this positive experience (eg, change your attitude or behavior)?</td>
<td></td>
</tr>
<tr>
<td>4. Can you tell us about an experience that turned out differently than you expected? What happened and how did you feel at the time?</td>
<td></td>
</tr>
<tr>
<td>5. Regarding this last experience where you wished things had turned out differently, did you or your doctor do anything to rectify the situation?</td>
<td></td>
</tr>
<tr>
<td>6. Did your participation in the study change your outlook on the way you manage your exacerbations?</td>
<td></td>
</tr>
<tr>
<td>7. What do you think would be the best way to integrate this type of long-term follow-up so that it addresses the aspects that are most important to you in the management of your exacerbations?</td>
<td></td>
</tr>
<tr>
<td>8. Is there anything else you wish to tell us about?</td>
<td></td>
</tr>
</tbody>
</table>

Textbox 3. Guide for the focus groups with care teams.

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. From the point of view of the health care team, what are the most important aspects in the management of patients’ respiratory exacerbations, particularly in their daily lives?</td>
<td></td>
</tr>
<tr>
<td>2. In your opinion, how have the proposed monitoring methods, including connected devices and patient education, addressed these priorities or with what limitations?</td>
<td></td>
</tr>
<tr>
<td>3. During this research project, what changes have you noticed in the way the team works or in its workload with regard to monitoring patients for the management of their exacerbations? Have you noticed a change in your relationship with the patients’ out-of-hospital physiotherapist?</td>
<td></td>
</tr>
<tr>
<td>4. What difficulties or bad experiences have you had in the process of managing patient exacerbations using connected devices?</td>
<td></td>
</tr>
<tr>
<td>5. Do you feel that you had positive experiences during this study with the management of patient exacerbations? How would you rate these experiences in relation to the most important aspects of the management of respiratory exacerbations?</td>
<td></td>
</tr>
<tr>
<td>6. In your opinion, should this type of long-term patient follow-up be included in the management of exacerbations or in other aspects of their management? If so, what would be the best way to integrate it and for which patients and with which objectives?</td>
<td></td>
</tr>
<tr>
<td>7. Is there anything else you wish to tell us about?</td>
<td></td>
</tr>
</tbody>
</table>

Analysis Framework

All the interviews were transcribed verbatim and subjected to a descriptive qualitative analysis. The analysis framework used was grounded theory [23].

Grounded theory is a qualitative research method with an inductive approach aimed at constructing a theory on a cultural, social, or psychological phenomenon by proceeding with the progressive and valid conceptual representation and mapping of qualitative empirical data [24]. In this study, the phenomenon explored was learning and empowerment in health management through the use of connected devices. Grounded theory is relevant as this phenomenon is currently sparsely studied. Studies on connected devices in patients with chronic conditions, and especially in patients with CF, are mostly intended to demonstrate the efficacy of the use of connected devices on various health outcomes. The theories mostly reported in the literature, such as digital behavior change interventions or the theory of reasoned action, are mainly focused on compliance with connected devices. However, the concept of empowerment includes other dimensions, such as understanding, the ability to decide, and self-assessment. Using grounded theory, we aimed to complete the current knowledge by eliciting the various dimensions of empowerment from the patient experiences with the use of connected devices for remote monitoring of their symptoms and by identifying elements that could enrich the theories in the field of remote monitoring.

According to the constructivist grounded theory method by Charmaz [25], which focuses on social processes or actions and the meaning of human actions, we adopted a social psychological approach to explore how and in which context individuals feel that connected devices have an impact on their learning to take care of themselves and on their empowerment. In grounded theory, verbatim transcripts are analyzed using codes to highlight what was stated by the participants in the study and derive meaning from it. We applied the standard steps of grounded theorizing. In initial coding, we generated as many ideas as possible inductively from the initial data. In focused coding, we relied on the most prevalent and important codes to select the central codes for analysis. In theoretical coding, we refined the final categories of the theory by connecting them to...
each other, thus allowing for the integration of the categories into a model and the construction of a theory on the phenomenon studied [26].

The grounded dimensional analysis of patients’ or parents’ and health care providers’ data was conducted by 2 researchers (MM and DPB) using NVivo (QSR International) taking into account their evolution over the course of the study and the various natures and production conditions of the collected material while constantly comparing the data within and across patients or parents and health care providers. A constant comparison between the verbatim transcripts of patients and health care professionals was carried out to bring out the invariant elements as “the essence of the phenomenon,” which elaborate “conceptual categories” remaining as close as possible to the lived realities of patients [27]. The 2 researchers who analyzed the verbatim transcripts were a psychologist and the parent of a child with CF, and both had PhDs in public health and great experience in qualitative research.

Analysis of the Educational Documents

The educational documents filled in by the physician and the patient during the educational session (educational workshop 2) were collected by the research team and reviewed globally but were not analyzed in connection with the patient interview. The aim was to understand which actions had been agreed upon between the patient and the physician when symptoms of a PEx were detected by the patient at home (central column) and whether they could resolve the PEx episode and prevent deterioration through their actions.

Ethical Considerations

The research project was submitted for evaluation by the Committee for the Protection of Persons designated randomly under conditions provided for in the Code of Public Health (Article L. 1123-14). The study was approved by the Committee for the Protection of Persons (CPP North West III) on June 10, 2017 (2017-A00723-50). Free and informed consent was obtained before any act related to research was undertaken.

Results

Population Interviewed and Dropout Rates During the Study

A total of 56% (20/36) of the study participants were interviewed. The population interviewed in relation to the population included in the study and who benefited from the different stages of the educational program (educational workshops 1-3) is listed by center in Table 1. The dropout rate at the end of the first phase of intensive data collection (3 months) was 25% (9/36). In total, 3% (1/36) of the patients died during the study. The death was unrelated to the study. A total of 67% (24/36) of the patients were educated in the first 2 workshops (educational workshops 1 and 2), allowing them to enter the routine monitoring phase using connected devices. Only 39% (14/36) of the patients attended the third educational workshop held at the midpoint of the routine monitoring phase using connected devices. At the end of the study, the nonresponse rate to interview solicitations compared with the number of patients who entered the routine monitoring phase was 25% (6/24). These results differed from one center to another. The gender, age, and geographic area characteristics of the patients interviewed (presented in Table 2) were similar to those of the entire study population. However, the patients interviewed had a higher level of education and employment rate than the entire study population.

A total of 12 health care providers from 7 hospitals participated in focus groups between May 2020 and February 2021 (Table 3).

Table 1. Number of patients interviewed per center (n=20).

<table>
<thead>
<tr>
<th>Patients included (n=36), n (%)</th>
<th>Patients educated, n (%)</th>
<th>Patients interviewed (n=20), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EW¹ 1 (n=30)</td>
<td>EW 2 (n=24)</td>
</tr>
<tr>
<td>Pediatric CFb centers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4 (11)</td>
<td>4 (13)</td>
</tr>
<tr>
<td>2</td>
<td>5 (14)</td>
<td>3 (10)</td>
</tr>
<tr>
<td>3</td>
<td>5 (14)</td>
<td>4 (13)</td>
</tr>
<tr>
<td>Adult CF centers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8 (22)</td>
<td>6 (20)</td>
</tr>
<tr>
<td>5</td>
<td>6 (17)</td>
<td>6 (20)</td>
</tr>
<tr>
<td>6</td>
<td>3 (8)</td>
<td>3 (10)</td>
</tr>
<tr>
<td>7</td>
<td>5 (14)</td>
<td>4 (13)</td>
</tr>
</tbody>
</table>

¹EW: educational workshop.
²CF: cystic fibrosis.
Table 2. Characteristics of the study participants (patients; n=20).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>Adolescents, n (%)</td>
<td>8 (40)</td>
</tr>
<tr>
<td>Adults, n (%)</td>
<td>12 (60)</td>
</tr>
<tr>
<td>Age of adolescents (years; n=8), median (SD)</td>
<td>14.5 (1.1)</td>
</tr>
<tr>
<td>Age of adults (years; n=12), median (SD)</td>
<td>29.6 (7.7)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>9 (45)</td>
</tr>
<tr>
<td>Female</td>
<td>11 (55)</td>
</tr>
<tr>
<td>Geographical area, n (%)</td>
<td></td>
</tr>
<tr>
<td>Living in a city</td>
<td>12 (60)</td>
</tr>
<tr>
<td>Living near a city</td>
<td>5 (25)</td>
</tr>
<tr>
<td>Living in the countryside</td>
<td>3 (15)</td>
</tr>
</tbody>
</table>

Table 3. Focus group participant characteristics.

<table>
<thead>
<tr>
<th>Hospital (n=20), n (%)</th>
<th>Date</th>
<th>MD^a (n=3), n (%)</th>
<th>Nurse (n=4), n (%)</th>
<th>Physiotherapist (n=3), n (%)</th>
<th>Other (n=2), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2 (10)</td>
<td>November 6, 2020</td>
<td>1 (33)</td>
<td>1 (25)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>2</td>
<td>3 (15)</td>
<td>February 5, 2021</td>
<td>1 (33)</td>
<td>1 (25)</td>
<td>1 (33)</td>
</tr>
<tr>
<td>3</td>
<td>3 (15)</td>
<td>November 9, 2020</td>
<td>1 (33)</td>
<td>1 (25)</td>
<td>1 (33)</td>
</tr>
<tr>
<td>4</td>
<td>3 (15)</td>
<td>October 11, 2020</td>
<td>0 (0)</td>
<td>2 (50)</td>
<td>1 (33)</td>
</tr>
<tr>
<td>5</td>
<td>4 (20)</td>
<td>December 9, 2020</td>
<td>1 (33)</td>
<td>1 (25)</td>
<td>1 (33)</td>
</tr>
<tr>
<td>6</td>
<td>3 (15)</td>
<td>June 23, 2020</td>
<td>1 (33)</td>
<td>1 (25)</td>
<td>1 (33)</td>
</tr>
<tr>
<td>7</td>
<td>2 (10)</td>
<td>May 19, 2020</td>
<td>1 (33)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

^aMD: doctor of medicine.

The forms completed by patients exiting the study reported technical difficulties with certain connected devices, in particular with the tablet computer provided to synchronize data before sending them to the research server. These difficulties were also reported by health care providers in the focus groups, who mentioned that a member of the team (nurse, physiotherapist, or clinical research associate) had spent a significant amount of time solving technical problems with the device suppliers, sometimes unsuccessfully. The reasons for dropping out of the study were multiple and are listed in Multimedia Appendix 3.

Results From the Educational Documents

The educational documents collected after educational workshop 2 show that the first column (“normal state of health, routine activities and treatments actually followed”) was filled with detailed information on the treatments and activities of the patient in their daily life, unlike the central column, which contained little information. The agreed upon actions in case of signs of exacerbation were mainly “increase physiotherapy” or “try to do more physical activity” and always “call or send a message to the center team.” The actions were aimed more at the diagnosis of the exacerbation by the physician, who then decided what the patient should do, than at the actions that the patients should take by themselves. Most of the physicians added the following comment—“They already know what to do”—meaning that they had not delegated new actions to the patients. One pediatrician decided to give conditional prescriptions of oral antibiotics to the parents after the educational session, thus delegating to them the decision to start the treatment and asking them to inform the team that they had started the treatment.

Descriptive Results From the Interviews

Stage 1: Initial Coding

A total of 12 codes emerged from the patients’ verbatim transcripts. In total, 10 codes emerged from the health care providers’ verbatim transcripts. The analysis allowed for the assignment of a name to each code that identified its area of interest (Table 4).
### Table 4. Codes and categories of transcripts from patient interviews and focus groups with health care providers.

<table>
<thead>
<tr>
<th>Category and codes from patient interview transcripts</th>
<th>Codes from transcripts of caregiver focus groups</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Category 1: patient empowerment</strong></td>
<td></td>
</tr>
<tr>
<td>- Learnings</td>
<td>- Learnings</td>
</tr>
<tr>
<td>• New knowledge mentioned by patients that helps them understand alerts and manage PEx&lt;sup&gt;a&lt;/sup&gt;</td>
<td>• Confirmation of the patients’ perceptions of symptoms using measurements; better understanding of their state of health at the first signs of exacerbation</td>
</tr>
<tr>
<td>- Empowerment</td>
<td></td>
</tr>
<tr>
<td>• Impression of being more capable of self-managing their treatments, their health, and their life projects</td>
<td>• Patient-physician relationship</td>
</tr>
<tr>
<td>- Loss of control</td>
<td></td>
</tr>
<tr>
<td>• Impression of being less capable of self-managing their care, health, and life projects</td>
<td>• Better understanding by the physician of the patient’s situation, their life circumstances, and their care; better understanding by the physician of the treatments carried out and the patient’s behavior in the event of exacerbation or in life in general</td>
</tr>
</tbody>
</table>

**Category 2: TTE<sup>b</sup>**

- Perceived usefulness
  - Needs expressed by patients to monitor PEx and expectations of the use of CDs<sup>c</sup> to help them self-monitor
- Perceived reliability
  - Patients’ level of trust in the reliability of the data collected by the devices during the study
- Negative experiences
  - Problems encountered using CDs; negative consequences described by patients

- Usefulness of monitoring using CDs
  - Depending on the patient’s health status (unstable or stabilized), on the caregiver’s previous experience with telemonitoring, and on the patient’s ability to use devices and keep them in good operating condition
- Technical reliability and accuracy of measurements
  - Checking the accuracy of the measurements taken using CDs in comparison with hospital standards and reliability over time
- Negative experiences
  - Problems encountered using CDs and negative consequences described by people—1 death that was not related to the study but that CDs did not prevent

**Category 3: use of technology by patients and health care providers**

- Conditions for a favorable use of CDs
  - Technical, human, and environmental conditions of CD use considered favorable for the optimal management of PEx
- Motivation
  - Personal and contextual factors that motivate patients to use CDs
- Hindrances
  - Personal and contextual factors negatively affecting the use of CDs
- Support from health care providers in the use of CDs
  - Support provided by health care providers in the use of data from CDs for the management of PEx that helps promote the use of CDs by patients
- Use of CDs
  - Modalities of CD use reported by patients

- Conditions of integration of the use of CDs into the organization of care
  - Technical, human, and organizational conditions for the health care team to integrate the support of the use of CDs by patients—resources and time needed for education and remote support of patients
- Factors of motivation in health care providers
  - Monitoring method that cannot be overlooked considering the current demographic increase in the number of adult patients; necessary monitoring method (using telecommunications) in case of a crisis (COVID-19)

N/A<sup>d</sup>
aPEx: pulmonary exacerbation.
bTTF: task-technology fit.
cCD: connected device.
dN/A: not applicable.

Stage 2: Focused Coding

Overview

Following the analysis of the initial codes, the codes from patients’ and health care providers’ verbatim transcripts were used to construct unified categories (Table 4) fed by the diversity of patients’ and health care providers’ perspectives. These categories are independent from one another and do not include the same codes. At this stage of analysis, some codes (“Patient recommendations” and “Implementation approach”) were set aside as they did not correspond to the modeling purpose. They will be considered later in the model. Each category was defined with a general title, a description, and detailed transcripts.

Category 1: Patient Empowerment

During the process of analysis, empowerment was defined as individual empowerment, characterized by the learning achieved during the intervention, the decisions and actions implemented by the patient for their care or health, and their sense of control over their health. Connected devices allow patients to access data on their health status daily to monitor episodes of PEx, prevent them, and adjust the course of action when they happen. They contribute to making some patients more autonomous in the early management of PEx by supporting their decision-making and ability to take action without seeing or contacting primary health care providers:

CDs allow for a better assessment of one’s health status, and to take better care of oneself. It helps to be more autonomous and to avoid waiting until we are very sick to go to the doctor’s. It also helps to complement one’s care with extra physiotherapy, more sports, things like that... [Adult patient]

The use of connected devices in the management of PEx not only allows for the adoption of preventive behaviors or better adherence to medical recommendations. Through their use of connected devices, patients also learn to manage their health with new data about themselves that confront the objective evolution of their health status with the way they feel and the effects of their lifestyle and their attitude toward their care. This process is characterized by the acquisition of new knowledge of one’s state of health, the validation of subjective perceptions, a better understanding of what happens on a physiological level, and focusing more attention on certain monitoring indicators. These learnings can be observed as early as during adolescence:

In the past, I probably used to desaturate without really realising it. I probably had headaches, but there you go...But now, I’m constantly stressed out, because I check my measurements pretty much all the time. [Adult patient]

Well, no. I can’t see the results of the measurements on the connected devices, and when the nurse called customer services, she was told that it was me who had the data anyway. But I don’t understand the data, and neither does she. So, perhaps we need to be taught how to interpret them better or to get clearer explanations in the alerts we receive. [Adult patient]

However, it has been found that measurements reflecting a deterioration can lead to higher stress levels and a loss of empowerment when no action plan has been put in place in advance with health care providers. A lack of patient education to support the understanding of data, including the meaning of alerts, can cause a feeling of helplessness in patients, especially if the caregiver also appears to be confused by the new monitoring method. Obtaining data on one’s health status on a near-continuous basis only enhances patient empowerment if the patient possesses the skills to interpret and act on the data. Similarly, connected devices lead patients to think almost constantly about their health despite not always being in the right mental state to do so:

Well, I found that the fact that I could make my own measurements ... allowed me to understand better...to be able to compare and to feel when I was not doing so well. And for example, I found it interesting when I thought I was doing less well but still had good results. I waited a little while to see if I should rely more on the results or more on how I felt, and I actually relied more on my results. And, yes, I thought it was good, because it’s mostly meant for prevention, and it helped me a few times. [Teenage patient]
improving patients’ state of health. Some health care providers emphasized the beneficial relational change fostered by the educational intervention that accompanied the implementation of connected devices, owing to which they communicated better with their patients. This allowed them to better understand their living conditions with the disease and how they cared for themselves daily. From their point of view, the data collected using connected devices increased patients’ level of information and of awareness of their condition. This gave the team the feeling of having a new tool to involve patients in the management of PEx and, thus, the capacity to influence the evolution of the disease:

*Behind the word “anticipation,” we mean they should know how to spot the early signs and manage to put things in place and then call us. They should know not to wait for one or two weeks before calling to tell us they haven’t been feeling well for two weeks.*

**Category 2: Adequacy of Technology (Combined With Education) to the Needs of Patients**

Often found in the literature as “task technology fit” (TTF) [28], this category includes aspects related to the reliability of the devices, the accuracy of the data measured (in comparison with a standard), the ease of access to the data, and how adequate the educational program is, all of which shape patients’ perception of how well this “technology” fits their monitoring needs. Patients expressed concern that the devices should accurately reflect their condition. The adequacy of the devices for monitoring purposes can be assessed based on several criteria over time as patients experience the use of connected devices. The first criterion is the perceived reliability of connected devices over time and the accuracy of the measurements compared with measurements taken at the hospital:

*When I took several measurements, I sometimes got very contradictory results. I sometimes wasn’t sure whether it was reliable.* [Adult patient]

The second criterion is the ergonomics and ease of use to connect to the tablet to access data and send them to the research server, which enables the sending of alert notifications:

*But the fact is that the spirometer...it does not save the results. So, I could get a good score at the beginning, but I tried again and because I coughed a little bit the result wasn’t so good, so I started again from the start, but it’s a bit difficult. Results should be saved automatically.* [Teenage patient]

The third criterion is the technical support provided for the implementation of the devices:

*What bothered me was that the curves on the graph—there were two curves—I never knew what they represented. And I even asked the nurse, and the nurse replied: “Indeed, it’s weird, what does it mean?” Even she searched for an explanation.* [Adult patient]

These aspects were supposed to be controlled in the context of interventional research, but some patients had a disappointing experience even though they were aware that they were participating in a pilot study that would sometimes involve “teething problems”:

*We told them that they were the first ones to go through all the steps and that everything was not necessarily perfectly set up for them...I told them that future patients would have fewer difficulties because we would manage to solve some things with them in the study. They tested the tools from beginning to the end and therefore experienced all the computer bugs.* [Health care provider]

The reliability of the devices used to monitor patients was mainly assessed by health care providers in comparison with the measurements taken using standard hospital equipment. This reliability was, from their point of view, guaranteed by the research context. Some patients felt that their health care providers did not have answers to the technical problems they faced:

*I received several emails from the CF centre telling me that they were not getting the data. But I assured them that I was sending the results. I managed to show them that I had uploaded the data...I went onto HealthMate as I was getting an update every Sunday by email for the Withings devices. So, I forwarded it to them, and in fact, they said that the data were loading, but not on the research server.* [Adult patient]

Patients’ interest in technology may vary according to the connected devices proposed, the need they feel to monitor certain health indicators, and the attractiveness of the device. Moreover, patients may not wish to use them for fear of being confronted with poor results on certain critical measurements for the patient (or for the physician):

*So, I found the sleep analysis option rather useful. Because I do sleep well at night, but I cough without realising it. I was either a little tired when I woke up in the morning, or even not at all tired, while it turned out that I had exacerbations at night. So, I could see that from two criteria: the first one was the decibel peak levels at night, and then the second one was when I didn’t have a restful night’s sleep. So, these were two rather useful criteria, I think. And then...yes, there also was a third one...It is my heart rate, which increased as soon as I coughed.* [Adult patient]

The integration of technology and patient education into the care process was seen as an additional workload by health care providers. Although dealing with technical problems took more time than expected for those in charge of the study (nurse or clinical research associate), physicians mainly mentioned the time spent on the patient education workshop (education workshop 2). Patient education undertaken by physicians in the adult patient care pathway is new for some adult centers, and those centers hope to benefit from a “return on investment”
from it in the future. From the point of view of the care team, taking measurements using connected devices adds to the time already spent by patients managing their disease daily:

For us, it takes time, but obviously, for the patients it represents a lot of time too. In patients’ daily lives, it clearly adds minutes to their basic treatment. In terms of the team’s workload, it obviously adds work, and the therapeutic education workshops linked to the protocol were particularly cumbersome. It’s a lot of work at the time, but it clearly is really beneficial for the future. [Health care provider]

Although the educational tool proposed in the study (educational workshop 2) was generally appreciated by adult patients, it may have seemed complicated to the adolescent audience although it was developed by a pediatric team and tested with several teenagers before releasing it to be used for research:

Therapeutic education went well too...The information pane (“React with CDs” tool) was really well done, and it allowed us to look into many habits that we didn’t have, well at least that I didn’t have. [Adult patient]
The dashboard was not bad, but super complicated to use for a teenager. There is too much stuff on it. And clearly, too much information on the same page. You can’t go straight to what you’re looking for...I mean, you really need to look for it. In that sense, I think this table needs to be more legible, because there was a lot of data on it. And reading a lot of data in a table with many columns, it’s...it’s not appealing. [Parent]

Personalized alert thresholds were set for each patient based on data collected during the first phase of the study following the statistical analysis (cumulative sum control chart). However, these alerts were rarely used by patients to manage their exacerbation episodes as reading measurement results alone allowed them to understand their health status or the lack of updates to thresholds rendered the alerts irrelevant:

At the end of the year, my FEV1 had increased by quite a bit, so when I started the new year with a new secondary infection, my FEV1 didn’t drop lower than the year before. As a result, I never received any alerts. So, I think in this case, we need to update the thresholds, because things can really fluctuate. [Adult patient]

Questions emerged among health care providers on the profile or profiles of patients for whom it is more relevant to introduce self-monitoring measures via connected devices. The inclusion criteria of the study targeted patients with good to moderate lung function (FEV1 >50%) so as to limit the risk of patients leaving the study because of lung transplantation, which is considered as soon as FEV1 decreases to <40%. Some physicians who followed adult patients believed that stabilized patients are good candidates for this follow-up through connected devices, whereas others pointed out that very unstable patients could benefit from this reactive warning system to manage decompensation. In such a critical situation, physicians emphasized the importance of systematically transmitting patient data to the center to help monitor the patients using alerts. Although most physician investigators wanted the study not to send patient data to the center as they felt that they did not have the resources to treat them, other physicians considered it not to be viable for patients who were critically ill. The fear of widening existing social inequalities in health was also mentioned by the care teams:

I think it is useful to integrate the use of such devices with severely ill patients who have frequent exacerbations, who are hospitalised...It can really have a positive impact by confirming the patient’s perception that they are not doing so well, and that they may need to begin an intravenous treatment. It can help patients and us, health care providers, for patients who are severely ill, by providing objective data on exacerbations.... But at the same time, we must not delude ourselves. It is with these severely ill patients that it will be more difficult to set up a monitoring process with CDs. Because they often are in complicated situations socially, psychologically, and so on. So, I don’t know whether it will really be possible with these patients. There are biases and inequalities that will remain true with CDs. Whereas patients who are already autonomous and stabilised will more easily appropriate the CDs. [Health care provider]

However, some patients want to maintain control over their data and make decisions themselves as they feared that connected device monitoring would increase the control of the care team:

I don’t need a doctor’s supervision to tell me to be careful and that today’s measurement was not good. Because on the contrary, I find it more worrying than anything else. But then, it depends on the CF centre. For example, some CF centres will use the measurements and overprescribe antibiotics, while others will want to see the patient in consultation.... It should be up to us, it’s our responsibility. [Adult patient]

Category 3: Device Use by Patients (and Health Care Providers)

In the context of this study, device use refers to the ways in which patients used connected devices, whether continuously or intermittently, which may have evolved during the course of the intervention according to factors linked to the patients’ life circumstances, what they experienced during the study, and the conditions of integration of the new monitoring process into the organization of the care team’s work. These uses reflect patients’ perceptions of the benefit-risk balance of the technology and its evolution during the study. Patients adapted the frequency of their connected device use to their need to self-monitor between quarterly visits to the center or, instead, to their need to “let go” slightly on disease management. This need for monitoring increases in periods such as the introduction or cessation of treatment, and it fluctuates depending on life circumstances (work), events related to the environment (high pollen count), or symptoms linked to the disease.
The following is an example of patients’ need for self-monitoring in between consultations at the hospital:

This allows us to watch the evolution of our data. The problem is that we go to the hospital once a month, or even every three months. So, we don’t have a regular follow-up as such. Whereas with these devices, for example, if I do a spirometry test once a week, I get a score every week, and I will check quite regularly, either it is effective or it is not. It’s complementary to my usual care and it could perhaps help patients be more autonomous. [Adult patient]

The following is an example of adopting connected device monitoring in specific situations or for particular diagnoses:

I am planning to get pregnant, and therefore, I think connected devices will be very useful during that time. Indeed, I may not be able to take all the treatments that I can usually take when I am not pregnant. So, I think the devices will be useful then and I also think I’ll be more conscientious in such circumstances. [Adult female patient]

The use of connected devices also depends on the way measures are integrated into the patient’s personal organization, also known as the routinization process, which, when compatible with their lifestyle, can alleviate the feeling of burden related to the use of devices and contribute to making the collected data more reliable. In the absence of a routine, the use of connected devices can also be taught through therapeutic education sessions and become part of a self-normative approach connected to the patient’s perceptions of their health status:

I do it when I have a quiet moment before leaving in the morning, before physiotherapy, and that’s it. I always tried to do it in the same conditions, so that it wouldn’t skew the data. [Adult patient]

In the particular case of adolescents monitored using connected devices, their use was regulated by the parents, which adds to the burden of preparation and control of certain treatments. The collaboration with an out-of-hospital physiotherapist in this monitoring was seen as a relief for the parent caregiver, and it emphasizes the importance (credibility) of the follow-up for the adolescent patient. The question of maturity related to patients’ age was raised regarding the implementation of monitoring using connected devices in adolescence. Conversely, a parent mentioned the help that these connected devices could bring for the empowerment of young patients. Additional notification functionalities inspired by other applications could also support their use of connected devices:

The greatest thing that could happen for kids would be that the watch sent them a notification if the scores were low and told them what to do. For example, we would set up some instructions onto the app, and as a result, they would receive notifications with the steps to follow on their watch. It would really make them autonomous then. Some apps allow the creation of a schedule and then send out notifications. Youngsters just have to look at their watch and it reminds them they have to bring a check on Monday at 10 AM to the school secretary to pay for the canteen. So, it doesn’t replace the parents, but it would relieve them of the task of always repeating things like a parrot, which causes a lot of conflicts in families. [Parent]

Sometimes, connected devices reactivated conflicts between parents and adolescents regarding the fear of addiction to the tablet for uses other than health monitoring or because they give parents access to data on the adolescent’s behavior:

There’s a very intrusive aspect to it. It feels quite overbearing for teenagers to know that they have lost 200 grams and that mum and dad want them to eat more to get the weight back on. Parental monitoring of sleep also creates conflict, and it was the case for almost all teens, with the parents saying: “You’re going to bed too late, that’s why you’re tired, it’s not healthy for you…” Some parents decided not to look at the data for that reason. [Health care provider]

The use of connected devices by patients is also determined by the interest and attention that health care teams pay to discussions on these data during consultations, phone calls, or teleconsultations with patients, which we will refer to as “patient support”:

So, I thought it was nice. It was really…I was sending screenshots of the saturation, well you know…and the FEV1. We had real conversations, and I found it interesting. It was more precise, less vague, the explanations I had to give…I had to give numbers, you know… [Parent]

Conversely, when health care providers fail to take into account information from connected devices in patient monitoring, it can make patients doubt the importance and usefulness of such data, which, in addition to the burden of taking measurements, can lead to a lack of interest in these devices:

We talked about it, but then, we didn’t focus the consultations on it at all…I expected there would be more guidance in terms of therapeutic education…we did it once about the information panel (“React with CDs” Tool), it took a very long time, it lasted almost two and a half hours. But I expected it would be that way during consultations, precisely to teach us to manage it ourselves…Sometimes, I wonder how it would be like if we had a chat every month, just for five minutes, just to ask me if things were going well, if there were any problems, or if I thought something was wrong. [Adult patient]

Beyond the use of connected devices for monitoring purposes between consultations, some patients suggested that this follow-up could allow them to space out their visits to the center, particularly when they lived far from the center or to limit the risk of contamination at the hospital:

Not on a regular basis, but sometimes when needed, to avoid going to the CF centre, because I live a little over an hour away. So, to avoid the journey, especially if I’m going to the hospital just to do a spirometry test to analyse FEV1, then yes, I might as
well do it with the device at home...it allows me to do the measurements myself. At the time of the consultation, we can either have a video call or talk over the phone, and then, we just give the results...in addition, we can extract our data, so we can even send them by email and the doctor can look at them beforehand. [Adult patient]

**Stage 3: Theoretical Coding**

The comparison of verbatim transcripts in the 3 categories revealed the relationships between them, as shown in Figure 2. These bidirectional relationships can be explained as follows. First, the TTF–empowerment relationship: this is reflected in patients’ trust in technology as a necessary condition to consolidate their learnings, which in turn strengthens their trust in the support they receive from technology. Second, the empowerment-use relationship: patients’ capacity to take action and their feeling of control over their health condition with the way the technology is used, which in turn strengthens their capacity to act on their health. This relationship is mediated by the support provided by the care team to help patients adjust the use of technology for their daily management of the disease and, therefore, improve the PEx diagnoses and the suitability of the prescription. Third, the TF-use relationship: the adequacy of the technology to patients’ needs influences its use by patients, reflecting the perceived advantage for patients of being monitored using devices compared with their “standard” follow-up at the CF center. Patients use devices more if they seem adapted to their needs and if they are reliable and easy to use. This relationship is mediated by the care team’s appropriation of the technology, which translates to their coordination of the remote monitoring, the use of real-life data in patient education, and them learning to master the use of devices for patient care.

*Figure 2.* Category modeling and mapping. CD: connected device.

Connecting all the results leads to theorization, the final stage of grounded theory analysis, which can be formulated as follows. The use of connected devices by patients results in an increase in their ability to take action over their health (empowerment) through the continuous adjustment of this use to their degree of autonomy, which influences and is also influenced by the conditions in which the technology is integrated into the organization of the care teams and the patient educational program. The motivation of patients with CF to use connected devices to prevent and manage PEx is dynamic. It depends on the patients’ priorities and specific concerns as well as triggers that will increase the feeling of usefulness related to the connected devices. The data provided by connected devices become a source of new knowledge (eg, about their disease and health) and capacities (eg, to prevent and manage) if a learning process to use them in daily life is implemented. This learning process can be supported by patient education.

This theory accounts for the conditions under which the implementation of connected devices for the management of PEx in patients living with CF can increase their capacity to act on their health.

**Discussion**

**Principal Findings**

Our study theorized the conditions that favor individual empowerment in patients living with CF in the management of PEx using connected devices as part of the MucoExocet study. This study took place in the context of a rare disease, proposed by the health care team to patients who could be interested in the technology. This theorization of the individual empowerment of patients through the use of connected devices is provisional, similar to any theory derived from the grounded theory approach, and remains subject to verification.
Our study places the concept of individual empowerment through the use of technology at the center of our research. Indeed, empowerment is an important mechanism of eHealth self-management, but validated assessment tools are rare [29]. In our theorizing, we viewed empowerment as the ability of individuals to identify and meet their own needs, solve their own problems, and mobilize the necessary resources through connected devices and education provided by health care providers to feel in control of the management of PE [10]. This leads us to consider the success of the implementation of connected devices from the perspective of the patient empowerment outcome, weighted by a typology of the intensity of connected device use by the patient deduced from our results. First, low use: patients doubt the feasibility and usefulness of this continuous monitoring; they are in favor of a “standard” follow-up at the CF center using data collected from the clinical examinations in situ. Second, advanced use: patients know that this monitoring can be useful in case symptoms appear or when a follow-up appointment at the CF center is not easily accessible. The use of connected devices remains optional and selective between the “standard” clinic visits at the center. Third, high use: patients experiment with the use of connected devices when in particular situations or because of particular behaviors, enabling them to consolidate or develop skills in relation to their health and the factors that influence it. This use is connected to their desire to improve their health and control its evolution and to the belief that the use of connected devices can support them in doing so. Depending on the patient’s situation (eg, developmental, emotional, and environmental), the level of motivation to engage in less or more intense connected device use will vary.

We suggest that empowerment, adherence to treatment, and quality of life be favored as primary outcomes of remote digital follow-up. In a recent study using a randomized trial that compared 2 groups of patients (using the tracker device against not using the device), Wildman et al [30] highlighted that the health improvement objectives were not achieved but that the intermediate objective of improving adherence to treatment was exceeded. These findings tend to confirm that, when assessing how effective the implementation of technology is with patients, the improvement of health indicators may not be the first outcome to be expected. This strengthens the case for a patient-based evidence evaluation approach.

In addition, the identification of “opposite cases” encountered in our study, for which patient empowerment was compromised, supports the theory stated—cases in which the devices were unreliable (TTF) or no action plan was defined in response to alerts or variations in measurements or cases of difficulties reaching health care providers (lack of support or difficulty integrating the use of technology into their organization) all led to lower levels of patient empowerment. Our study questions eHealth-backed education models, for which data are currently scarce. Following the work of Greenhalgh et al [31], our results highlight the interaction among the patient, the device, and the organizational and social system as the cornerstone of the learning process in patients. This interdependence underscores the systemic approach to connected device implementation, wherein connected device adoption and use and the positive experience with them cannot be attributed to the patient’s lack of motivation alone. Indeed, connected devices introduce a technopedagogical transformation among health care providers, which pushes them to rethink organizational and educational activities to support a new relationship with patients.

This study shows that connected devices may have enabled health care providers to gain a new understanding of patients thanks to the quality and novelty of the information obtained via connected devices. In this sense, connected devices could help bridge the gap that is sometimes observed between theoretical models based on medicine that is “centered on the person and their family” and the practice of care that lacks understanding of patients’ experiences in daily life [32]. Health care providers are made to understand the daily lives of patients living with a chronic disease in physical, psychological, and social terms, thereby creating a more symmetrical relationship of information sharing [33]. Our study shows that caregiver-patient interactions are modified by the introduction of connected devices. They are enriched by a new outlook on patients’ daily lives mediated by technology, which leads to a new understanding by health care providers.

In addition, this study confirmed that the implementation of connected devices should be considered based on patients’ health goals and not simply focused on education on the device [34]. Patient empowerment depends on the connected devices’ capacity to meet the needs of patients’ health project. Patients then enter a learning process supported by the connected device and with educational support from health care providers structured in 4 phases, as described by Almalki et al [35]: identification of an area of interest (the patient is focused on a specific health goal that requires the collection of data about themselves), personal analysis (analyzing one’s behavior in light of the objective data collected), self-experimentation (structuring a reasoning based on the trends identified in support of the experimentation carried out with the connected device), and activation (confirmation of the hypotheses made during the experimentation phase and development of personal knowledge). This process must be structured and accompanied to unfold properly.

Although a recent review of the literature [36] on the use of mobile devices (phones, patient monitoring devices, digital assistants, and other wireless devices) by patients with CF has shown medical, psychological, and behavioral benefits as well as benefits in terms of level of satisfaction with care, the psychological aspect has thus far received little attention, as is the case with the educational dimension of technological devices. In adolescence, although the disease significantly influences the development of one’s body image and self-concept [37], the integration of new technologies into self-care leads to a new understanding of oneself and, therefore, to a potentially modified relationship with one’s body, health, and illness. This process is an integral part of the use of new technologies. Therefore, it is a potential topic for future research, which is necessary to understand the use of new technologies in care and their effects on people.
Limitations of the Study
This pilot study was based on an interventional research protocol. On the one hand, this protocol was implemented differently depending on the centers and the devices selected for the research—the teams applied the educational program differently, the elaboration of self-management action plans in the event of exacerbations was done differently (educational workshop 2), or the midterm review session of the routine follow-up of the patient was different (educational workshop 3). These differences in the implementation of the protocol were noted when collecting the experience of patients and data from the focus groups; they contributed to enriching the definitions of the categories and the relationships between them. In contrast, within the framework of this research protocol, we could not modify the tools that proved to be unreliable (which had been selected in 2016 via a market analysis while planning for the study), adjust the formatting of the data, or more generally adapt the intervention according to the results collected throughout the early phases of the study. Thus, having a protocol that is too fixed is probably a mistake to avoid in health technology research if we wish to adapt the intervention during its implementation to explore the best way of using health technology. In the context of the intervention, the choice of connected devices and the setup chosen did not allow patients to access a dashboard displaying the data collected by all connected devices. Some more motivated patients created their own dashboards separately. Furthermore, the pulmonologists in charge of patient follow-up were not always involved in the study, and this dichotomy made routine monitoring more complicated for patients. Similarly, when the clinical research associate in charge of the study was not the patient’s coordinating nurse, the latter was unable to answer patients’ questions during consultations or phone calls. Eventually, the study included people interested in technology, which could have biased the results based on the experience of using technology.

If this had been a descriptive pilot study, a quality improvement approach would have allowed for adjustments and improvements to the intervention over the course of the study and would have been directly driven by the care team. This format has been used to introduce connected devices into the patient care process for CF in the United States [38], which enabled patients to be equipped and monitored remotely when CF centers were closed during the COVID-19 pandemic by using a connected spirometer coupled with teleconsultation. The quality approach allowed for the evaluation of the results during the course of implementation, and adjustments were made to the intervention to improve its impact. The results were convincing:

In March 2020, the beginning of the pandemic, 37% (49/131) of patients owned a HS (home spirometer) and around 50% (92/20) of patients seen via telemedicine performed spirometry at home. By September 2020, 97% (127/131) of adult patients at UVA owned a HS, and by October 2020, 96% (2425) of patients provided spirometry results during their telemedicine encounters.

Prospects for Transferability
Assessing how transferable the theory could be outside the context of its development would require studying the introduction of connected devices in other circumstances: with patients living with different diseases, using different devices, or with a different organization of care.

Two opposite contexts could be studied in terms of patient empowerment through technology: (1) a context of patient dependence on self-regulated or caregiver-driven technology, whether it is telemonitoring, implantable devices for which the use is predetermined (dependence on technology and on health care providers making the care decision in the event of an alert or emergency), or protocolized treatment with little margin for adaptation or action because of side effects (eg, protocol dependence in cancer treatments); and (2) a context of patient-developed technologies [39] made available to patients living with the same condition in open source, as is the case with type 1 diabetes mellitus (T1DM). T1DM has the highest degree of patient empowerment and has recently led to the publication of an international consensus for the guidance of professionals caring for patients who use such devices. The case of T1DM is also interesting as research was conducted on the transition of patients from devices that allow for the management of glycemia and insulin delivery in a semiautomated way to a closed-loop insulin delivery system, which is designed to “free” the patient from self-management by automating the process of insulin delivery. However, this specific case might also lead patients to feel that they lose control over their glucose levels before they take back control over some other parameters of the automated process.

Contribution to an Extended Theory of Empowerment From Remote Monitoring for Health Symptom Tracking
A recent publication by White et al [40] reports on a systematic review to help define engagement with remote monitoring for health symptom tracking (RMT) and how to measure it. Engagement is seen as a mediating factor that eventually explains the impact of RMT on patient health outcomes. Their analysis is of most interest to our own work and shows that concepts still need to be clarified in the context of RMT. They propose a definition of engagement through a remote monitoring protocol (dropouts), objective engagement, subjective engagement, and interactions between objective and subjective engagement. Although objective engagement (with remote monitoring itself, with symptom tracking compliance, and with app use of statistics) is clearly measurable, subjective engagement appears to gather a wide range of concepts, some of them from the technology acceptance model literature (usability, TTF, satisfaction with the technology, utility for symptom management, ease of use, and intention for future use; Davis [41] revised by Venkatesh et al [28] and Chang et al [42]).

In a further extended theory, we would rather build on certain determinants of the technology acceptance model and distinguish them from the concept of patient engagement. These determinants leading to the “behavioral intention of use” would be the personal characteristics (age and sex, expectations, social influence, hedonic motivations, and previous experiences with
information and communication technologies), the facilitating conditions over time, TTF (over time as technologies are continuously refined), and the mediating factors (perceived ease of use and perceived usefulness). We would propose to include the "engagement with the research protocol" by White et al [40] as a determinant, renamed as "conditions for the RMT introduction/intervention" (either research or routine care or self-care). Our study aimed to add elements to modulate the "behavior use" in the RMT context, which is not explained by the previous theories and not necessarily consistent with the "behavioral intention of use." From our study, these elements could refer to patient empowerment, such as their learnings about their own body, their trust in the technology, and the relationship and support they receive from their care team. We agree with the conclusion of White et al [40] to explore the RMT field in its own right as separate from Digital Behavior Change Interventions or general eHealth literature.

Conclusions

Our study allowed us to propose a theory on individual patient empowerment through the use of connected devices based on patients’ and health care providers’ experiences in the context of an interventional pilot study. This theory needs to be validated with a larger sample and verified in the context of different diseases, different devices, and a different organization of care. It implies that, if the empowerment of patients with chronic diseases is indeed a desirable goal for all parties involved (patients, health care providers, and the health care system), the necessary conditions for the successful implementation of connected devices cannot be looked at separately for each party (health care providers, patients, and health care system). On the contrary, these conditions must be adjusted to the overall collaboration among these stakeholders, who cooperate toward patient empowerment. Only if all these conditions are met can patient empowerment be the outcome of the use of technology.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The "React to PEx" educational tool.

Multimedia Appendix 2
Exit interview questionnaire.

Multimedia Appendix 3
Reasons for leaving the study.

References


Abbreviations

CF: cystic fibrosis
EQUATOR: Enhancing the Quality and Transparency of Health Research
FEV1: forced expiratory volume in 1 second
PEx: pulmonary exacerbation
RMT: remote monitoring for health symptom tracking
T1DM: type 1 diabetes mellitus
TTF: task technology fit

©Maxime Morsa, Amélie Perrin, Valérie David, Gilles Rault, Enora Le Roux, Corinne Alberti, Rémi Gagnayre, Dominique Pougheon Bertrand. Originally published in JMIR Formative Research (https://formative.jmir.org), 23.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Culturally Adapting the World Health Organization Digital Intervention for Family Caregivers of People With Dementia (iSupport): Community-Based Participatory Approach

Anna Messina¹, PhD; Rebecca Amati¹, PhD; Anna Maria Annoni¹, MS; Beatrice Bano¹, MS; Emiliano Albanese¹, MPH, MD, PhD; Maddalena Fiordelli¹, PhD

Institute of Public Health, Faculty of Biomedical Sciences, Università della Svizzera italiana, Lugano, Switzerland

Corresponding Author:
Anna Messina, PhD
Institute of Public Health
Faculty of Biomedical Sciences
Università della Svizzera italiana
Via Buffi 13
Lugano, 6900
Switzerland
Phone: 41 0782104055
Email: anna.messina@usi.ch

Abstract

Background: Informal caregivers of people with dementia are at high risk of developing mental and physical distress because of the intensity of the care provided. iSupport is an evidence-based digital program developed by the World Health Organization to provide education and support for the informal everyday care of people living with dementia.

Objective: Our study aims to describe in detail the cultural adaptation process of iSupport in Switzerland. We specifically focused on the participatory strategies we used to design a culturally adapted, Swiss version of iSupport that informed the development of the desktop version, mobile app, and printed manual.

Methods: We used a mixed methods design, with a community-based participatory approach. The adaptation of iSupport followed the World Health Organization adaptation guidelines and was developed in 4 phases: content translation, linguistic and cultural revision by the members of the community advisory board, validation with formal and informal caregivers, and refinement and final adaptation.

Results: The findings from each phase showed and consolidated the adjustments needed for a culturally adapted, Swiss version of iSupport. We collected feedback and implemented changes related to the following areas: language register and expressions (eg, from “lesson” to “chapter” and from “suffering from” dementia to “affected by” dementia), resources (hyperlinks to local resources for dementia), contents (eg, from general nonfamiliar scenarios to local and verisimilar examples), graphics (eg, from generalized illustrations of objects to human illustrations), and extra features (eg, a glossary, a forum session, and a read-aloud option, as well as a navigation survey).

Conclusions: Our study provides evidence on how to culturally adapt a digital program for informal caregivers of people living with dementia. Our results suggest that adopting a community-based participatory approach and collecting lived experiences from the final users and stakeholders is crucial to meet local needs and to inform the further development, testing, and implementation of digital interventions in a specific cultural context.

(JMIR Form Res 2024;8:e46941) doi:10.2196/46941

KEYWORDS
informal caregivers; iSupport; dementia; digital interventions; mHealth; community-based participatory research; community; caregiver; mental distress; physical distress; support; development
**Introduction**

**Background**
Approximately 55 million people are currently living with dementia worldwide [1]. Switzerland accounts for >150,000 cases of dementia, with an expected doubling by 2050 [2]. In Switzerland, as in most countries, the majority of people living with dementia live at home assisted by an informal caregiver, who is usually a family member who provides daily support and coordinates care delivery [3]. There are positive outcomes that may be associated with the caring role, such as the perception of a better relationship and closeness with the care recipient [4]. Nonetheless, the increasing complexity of taking care of a person affected by dementia exposes informal caregivers to psychological distress and increases the risk of loneliness and developing symptoms of anxiety and depression [5,6]. Caregivers’ psychological distress is also associated with a lower quality of care provided [7] and with the worsening of behavioral and psychological symptoms in the care recipients [8].

Providing guidance and support to informal caregivers is one of the priority areas identified by the World Health Organization (WHO) to reduce the global impact of dementia and to improve the quality of life of caregivers and their families [9]. In the last 2 decades, digital educational and psychosocial interventions for caregivers have bloomed [10,11]. Internet-based interventions are more easily accessible [12,13] and adaptable to the time and geographic constraints of caregivers [14]. Some reviews suggest that multiple components of digital interventions can contribute to reducing the burden and improving the quality of care and be even more beneficial if tailored to caregivers’ specific needs and contexts [10,15,16]. The active involvement of the final users and relevant stakeholders in the design and local adaptation as well as the testing and piloting of interventions is crucial for need-centered interventions in terms of their uptake, integration, and scalability at the community level [17,18]. However, more evidence is required to understand the most effective methods and strategies needed to involve participants in the design and adaptation of digital interventions [19,20].

iSupport is an evidence-based digital training intervention developed by the WHO to provide support and education to informal caregivers of people with dementia [21]. The original program consists of 22 thematic lessons distributed across 5 modules (Figure 1). Each lesson covers a specific topic associated with care that ranges from the daily assistance of the care recipient (eg, toileting, personal care, and nutrition) to the self-care of the carer (eg, reducing stress and involving others in care duties). All lessons include theoretical and informative sections and case scenarios with interactive multiple-choice questions. The WHO provides guidelines to culturally adapt iSupport contents to the local language, culture, and context before implementation [22]. The reporting of adaptation processes of complex interventions is limited but extremely important [23]. Knowledge exchange of methodologies and approaches as well as evidence on barriers and facilitators to local adaptation are crucial preliminary steps to inform the implementation of interventions and their mid- to long-term uptake and sustainability [24].
Objectives
This study aims to describe in detail the cultural adaptation process of iSupport in Switzerland. We specifically focused on the participatory strategies we used to design a culturally adapted, Swiss version of iSupport that informed the development of the desktop version, mobile app, and printed manual. Our purpose is to inform the implementation of not only iSupport but also other complex health interventions, specifically in the context of the cultural adaptation process.

Methods
Study Setting
The study took place in the Italian-speaking part of southern Switzerland, namely the canton of Ticino.

Ethical Considerations
Before initiating the study, we sought ethics oversight by submitting our project to the cantonal ethics committee, and we obtained a waiver of ethics approval and official authorization to proceed with the study (project ID: 2020-02030 / CE 3731).

Specifically, the ethics committee determined that our project did not fall within the scope of the Swiss Federal Human Research Act [25], thereby granting us permission to proceed.

Theoretical Approach
The overall process of culturally adapting iSupport in Switzerland was based on principles from the community-based participatory research (CBPR) framework, which can be defined as "an approach to research that involves collective, reflective and systematic inquiry in which researchers and community stakeholders engage as equal partners in all steps of the research process” [26]. In intervention research, adopting CBPR has the advantage of facilitating knowledge exchange between the community and the researchers, reducing potential power imbalances, and increasing the likelihood of intervention uptake and success [27].

We based the specific phases and procedures of the adaptation process on the WHO iSupport adaptation guidelines [22], which, in turn, are based on the ecological validity framework proposed by Bernal et al [28] that is widely used for developing culturally sensitive interventions and strengthening their ecological validity [29-31].
Study Procedure

The Community Advisory Board

At the outset, we established a community advisory board (CAB) comprising community members and organization representatives who shared a common identity, geography, language, culture, and other values and principles [32].

We identified potential members of the CAB through a structured stakeholder analysis and mapping that accounted for the different levels of power, importance, and interest of the stakeholders in the project. We included representatives of the project’s funding agencies and other collaborating partners, caregivers of people with dementia, and members of the IT service in charge of developing the iSupport web platform and app. Once consensus among researchers was reached, we contacted and informed the identified members via email using a brief description of the project, the scope of the CAB, and their expected roles and responsibilities.

In the context of iSupport adaptation, the specific roles of establishing a CAB were to (1) help researchers to identify the needs and legitimate interests as well as the expectations of the different stakeholders and the final users and (2) inform the development of the intervention throughout a purposely co-designed process.

The adaptation process of iSupport consisted of four phases: (1) content translation, (2) linguistic and cultural revision, (3) validation with formal and informal caregivers, and (4) refinement and final adaptation. Each phase was based on, and adapted from, the WHO guidelines. Any change or proposed addition was discussed with, and approved by, the WHO. The members of the CAB were constantly informed and updated on the progress of the study. A flowchart of the phases is summarized in Figure 2.

Figure 2. Flowchart of the adaptation process of iSupport in Switzerland. CAB: community advisory board; WHO: World Health Organization.

Phase 1: Content Translation

The first step in the cultural adaptation of iSupport was the translation of the contents (approximately 60,000 words) from English, the original language of the program, into Italian, the local language in southern Switzerland. The process started in May 2020 and ended in August 2020. According to the WHO guidelines, the translation should be accurate while recognizing the local culture and its people. In line with this, we conducted a preliminary adaptation of culturally sensitive terms, including (1) personal names of the characters used in the case studies, (2) available information materials and local services, and (3) reference to cultural habits and leisure activities in the region [22].

One member of the research team fluent in English, AM, a psychologist with previous expertise in the dementia field, translated the original contents of the iSupport program into Italian. Subsequently, a senior member of the team, MF, with expertise in the field of health communication, checked the translations and proposed changes and modifications. All disagreements or doubts about the translation of sensitive terms
and expressions were documented and discussed within the research team in meetings until a consensus was reached. We sought the support of an external professional translator to resolve some specific language locations and terms.

Throughout the process, translators applied the international standards and available dementia guidelines to avoid stigmatizing expressions and to use language that promotes the inclusion and dignity of people living with dementia and their carers [33]. During this phase, we did not apply any changes to the meanings of the original structure of the iSupport program, including case studies or activities. All translations were copied into secure Microsoft Word files and stored in a dedicated Microsoft Teams workspace to optimize efficiency.

**Phase 2: Linguistic and Cultural Revision**

In September 2020, the first CAB meeting took place with the main goals of introducing the members of the CAB to the iSupport program and the research team and clarifying their roles and involvement throughout the research process. During the meeting, we answered all questions and proposed an interactive activity where participants were asked to provide the translation from English into Italian of a selection of sensitive terms and expressions used in iSupport that were noted by researchers during phase 1. At the end of the introductory meeting, participants were asked to sign a letter of intent that summarized the functioning of the CAB and their role and commitment as members of the local iSupport CAB. We explicitly specified the structure and definition of the CAB; goals, roles, and responsibilities (of both CAB members and the research team); and duration (Multimedia Appendix 1). All 9 invited participants agreed to join the iSupport CAB: 4 (44%) were informal caregivers, and the remaining 5 (56%) included representatives of the government (1/5, 20%), the local Alzheimer association (1/5, 20%), a health care service provider (1/5, 20%), the IT service (1/5, 20%), and the University of Applied Sciences and Arts of Southern Switzerland (1/5, 20%).

In October 2020, we shared the translated contents of iSupport with the members of the CAB and asked them to evaluate, and provide feedback on, each chapter and module of the program by the end of December 2020.

On the basis of the work of previous adaptations of iSupport [34], participants were asked to carefully go through the 23 thematic lessons and assess the translation and preliminary adaptation of iSupport considering six main parameters: (1) familiarity, (2) sensitivity, (3) comprehensibility, (4) precision, (5) cultural adequacy, and (6) overall evaluation. In addition, they were asked to assess the extent to which (1) the terms used were familiar to the target group (eg, the use of idioms and figures of speech), (2) the language used respected and promoted the dignity of people living with dementia and their carers (eg, the use of stigmatizing terms), (3) the contents were intelligible and easy to understand (eg, minimal use of technical jargon), (4) the contents were presented in an accurate way (eg, they were in accordance with the facts, and there were no mistakes), (5) the contents were appropriate and reflected the experiences of local people (eg, case studies), and (6) the content of each chapter was overall culturally appropriate. At the end of each chapter, participants were asked to fill out a digital survey via Research Electronic Data Capture (REDCap; Vanderbilt University) [35,36] to evaluate each of the aforementioned parameters using a Likert scale ranging from 1=requiring an extensive revision to 4=no additional revision needed. We also invited participants to provide additional comments about individual chapters through a dedicated open-ended question in the survey or to provide free feedback on the overall program via email. The survey was specifically designed for the purpose of this phase and was based on the work of Teles et al [34] to evaluate the cultural adequacy of the contents, as recommended by the WHO adaptation guidelines [22].

After the data collection period, AM and BB (a research assistant with a degree in psychology and health communication) performed a descriptive analysis of the quantitative data and a thematic analysis of the qualitative data. For the quantitative analysis, we used SPSS statistical software (version 25.0; IBM Corp) [37] for Windows to compute mean scores for each program module and survey parameters. For the qualitative analysis, we performed a thematic content analysis of open comments [38]. The maintenance of scientific rigor was ensured through regular meetings among research team members, particularly involving MF and RA, both experts in qualitative research methods.

In January 2021, the main findings of this phase were summarized in a report shared across, and approved by, all CAB members.

**Phase 3: Validation With Formal and Informal Caregivers**

We adopted a qualitative descriptive design, and we used focus groups (FGs) as a data collection method [39]. Between June 2021 and August 2021, we conducted FGs with formal and informal caregivers to explore their attitudes toward, and impressions of, the adapted version of iSupport. We prompted and collected suggestions for improvement, as also recommended by the WHO guidelines. In addition, we decided to expand our inquiry to caregivers’ attitudes toward support measures and help-seeking behaviors, which we have previously reported in detail elsewhere [40].

From April to May 2021, we crafted an invitation letter and a flyer presenting the project, the main purpose of the FGs, the eligibility criteria, and contact information. We disseminated these materials in a local newspaper, to members of the CAB and their associations and institutions, to daycare centers for people with dementia, and to participants of other ongoing research projects who had consented to be informed about further research activities.

Eligibility criteria for both formal and informal caregivers included (1) having (at present or in the past) experience in caring for a person living with dementia, (2) being fluent in Italian, and (3) living in the canton of Ticino. Caregivers who met the inclusion criteria could contact us via email or telephone. Those who contacted us were given an overview of the iSupport program, with excerpts from the same translated material used in phase 2, and the informed consent form (Multimedia Appendix 2). The FGs, which lasted approximately 2 hours, were audio recorded and took place either digitally via the Zoom
platform (Zoom Video Communications, Inc) or in person at the Università della Svizzera italiana in Lugano, Ticino. AM moderated all FGs, with the supervision of MF and RA. The discussions were transcribed verbatim and pseudonymized by EB, an independent research assistant. AM, RA, and MF performed a thematic content analysis to identify key themes [38]. Initially, the researchers familiarized themselves with the data through repeated reading of notes and transcripts to get an idea of the overall meaning and begin discerning key themes. Subsequently, each researcher independently identified codes within each FG (vertical analysis) and across the whole data set (horizontal analysis) to uncover variations and patterns within the data. Themes were progressively refined and consolidated through discussion in weekly meetings over 4 months (from November 2021 to February 2022) and until a consensus was reached. Data management and coding processing were facilitated by NVivo 12 software (Lumivero) [41]. Additional methodological details have been previously reported [40].

**Phase 4: Refinement and Final Adaptation**

All data collected during phases 2 and 3 were collated to generate a set of proposed changes and adaptations to the iSupport program. AM and BB familiarized themselves with the data and differentiated the feedback between cross-cutting and general comments and specific chapter–related comments and arranged them across 5 dimensions: language, resources, contents, graphics, and extra features. Each comment was then discussed between AM and BB and categorized as (1) rejected/not applicable, (2) possibly applicable, and (3) applicable. The categorization was based on the number and contents of suggestions received as well as in accordance with the WHO adaptation guidelines [22]. The feedback data that were considered possibly applicable and applicable were then discussed with the other members of the team to agree on their potential implementation.

Subsequently, all specific and applicable comments were charted using the iSupport WHO adaptation spreadsheet (Multimedia Appendix 3). All proposals of change were then supported by a rationale and by the source of the proposal: the research team (preliminary adaptation during phase 1), the members of the CAB (results from the linguistic and cultural adaptation during phase 2), and FG participants (data collected during phase 3). Attached to the adaptation spreadsheets, we also sent a list of general suggestions and feedback. The material was sent for revision to the authors of the WHO iSupport program in October 2021, and the results of their final fidelity check were received in January 2022. Subsequently, the local research team implemented all approved changes and uploaded the new adapted contents on the beta version of the iSupport Swiss web platform.

**Results**

A detailed description of all final adaptations made to the original iSupport program, resulting from the 4 phases of the adaptation process, is presented in Multimedia Appendix 4. The results of phases 2 and 3 are summarized in the subsections that follow.

**Adaptations From Phase 2: Linguistic and Cultural Revision**

**Overview**

All 9 members of the iSupport CAB revised ≥1 modules of iSupport and provided feedback, with each module revised by at least 1 CAB member. Module 5, which dealt with behavior changes, was the most revised and received the most comments (7/9, 78%). Descriptive analysis showed that all modules were generally positively evaluated with an overall mean evaluation score of 3.72 (SD 0.00) out of 4. Mean scores of the parameters across all modules ranged from 3 (SD 0.00) to 3.89 (SD 0.47) for sensitivity, from 3 (SD 0.00) to 4 (SD 0.00) for familiarity, from 3.93 (SD 0.26) to 4 (SD 0.00) for comprehensibility, from 3.67 (SD 0.58) to 4 (SD 0.00) for the accuracy of the information, and from 3.59 (SD 0.62) to 4 (SD 0.00) for cultural adequacy (Table 1).

We conducted a qualitative thematic analysis of the open comments and identified 7 potential areas for the improvement of iSupport (for more details, refer to Multimedia Appendix 4).

**Table 1.** Mean scores of the 6 parameters for the linguistic and cultural revision of iSupport.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Module 1, mean (SD)</th>
<th>Module 2, mean (SD)</th>
<th>Module 3, mean (SD)</th>
<th>Module 4, mean (SD)</th>
<th>Module 5, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>3.60 (0.55)</td>
<td>3.59 (0.62)</td>
<td>3.00 (0.00)</td>
<td>3.87 (0.52)</td>
<td>3.89 (0.47)</td>
</tr>
<tr>
<td>Familiarity</td>
<td>3.60 (0.55)</td>
<td>3.65 (0.61)</td>
<td>3.00 (0.00)</td>
<td>3.67 (0.62)</td>
<td>4.00 (0.00)</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>4.00 (0.00)</td>
<td>3.94 (0.24)</td>
<td>4.00 (0.00)</td>
<td>3.93 (0.26)</td>
<td>4.00 (0.00)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>3.75 (0.50)</td>
<td>3.76 (0.44)</td>
<td>3.67 (0.58)</td>
<td>3.73 (0.70)</td>
<td>4.00 (0.00)</td>
</tr>
<tr>
<td>Cultural adequacy</td>
<td>3.60 (0.55)</td>
<td>3.59 (0.62)</td>
<td>4.00 (0.00)</td>
<td>3.80 (0.56)</td>
<td>3.89 (0.47)</td>
</tr>
<tr>
<td>Overall assessment</td>
<td>3.20 (0.45)</td>
<td>3.65 (0.49)</td>
<td>3.00 (0.00)</td>
<td>3.80 (0.41)</td>
<td>4.00 (0.00)</td>
</tr>
</tbody>
</table>

**Familiarity With Terms and Expressions**

The unfamiliarity with the terms referred especially to some expressions that were largely used throughout the text to designate caregivers and health care workers or dementia health and social care facilities, and the comments by the CAB members allowed us to improve the translations; for instance, “informal carers” and “paid in-home helpers” were newly translated using local terms that were easily identifiable and familiar to participants (eg, “informal carer” was replaced with “familiare curante,” which literally means “family carer”). Importantly, the term is also used at an institutional level [42] to refer to people who take care of a loved one (relative or friend) affected by a chronic disease.
Sensitivity of the Language

Participants also suggested improving the sensitivity of the language by removing expressions potentially stigmatizing such as “suffering from dementia,” which was replaced with “affected by dementia” (affetto da demenza), or “show compassion,” which was replaced with “show comprehension” (mostrare comprensione).

Scientific Accuracy of the Terms

The jargon used was generally perceived as comprehensible and easy to understand. However, some of the participants (2/9, 22%), especially those working in the field of dementia, reported the need to use scientific terms to improve the accuracy of the language and ultimately the users’ literacy. As a result, the expression “memory loss” was, for example, replaced with “memory impairment” (difficoltà di memoria), and “helpful/unhelpful thoughts” was replaced with “functional or dysfunctional thoughts” (pensieri funzionali e disfunzionali).

Educational Approach

The educational approach referred to the use of terms considered scholastic and potentially belittling by participants, such as “lesson” and “learn,” which were replaced with “chapter” (capitolo) and “know more about” (conoscere di più), respectively.

Use of English

Some English terms that were retained during the content translation because they are normally used in spoken Italian were translated into Italian, including “focus” (obiettivo) and “relax” (rilassa).

Use of Numbers

To make the reading smoother, some of the participants (3/9, 30%) suggested replacing numbers with sentences (eg, from “20%-30%” to “approximately one-third”; circa un terzo).

Language Register

Finally, almost all participants (8/9, 89%) found that the tone and prose were at times informal or even childish; therefore, for example, the original sentence at the end of each chapter “you finished the lesson, well done” was replaced with “you finished the chapter, let’s go to the next!” (hai completato il capitolo, passa al successivo!).

Adaptation From Phase 3: Validation With Formal and Informal Caregivers

Overview

Between May 2021 and August 2021, we conducted 6 FGs: 1 (17%) with formal caregivers and 5 (83%) with informal caregivers. Most of the participants (16/19, 84%) were female, and the formal caregivers (6/19, 32%) had longer years of caring experience than the informal caregivers (13/19, 68%). The main characteristics of the caregivers are reported in detail in Tables 2 and 3.

<table>
<thead>
<tr>
<th>ID</th>
<th>Sex</th>
<th>Age (y)</th>
<th>Employment status</th>
<th>Years of professional caring experience</th>
<th>Years of personal caring experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Female</td>
<td>52</td>
<td>Housewife or retired</td>
<td>&gt;10</td>
<td>&gt;10</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
<td>54</td>
<td>Housewife or retired</td>
<td>&gt;10</td>
<td>&gt;10</td>
</tr>
<tr>
<td>3</td>
<td>Female</td>
<td>28</td>
<td>Housewife or retired</td>
<td>6-10</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>Female</td>
<td>45</td>
<td>Employed</td>
<td>&lt;1</td>
<td>&lt;1</td>
</tr>
<tr>
<td>5</td>
<td>Female</td>
<td>29</td>
<td>Housewife or retired</td>
<td>1-2</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>Female</td>
<td>59</td>
<td>Employed</td>
<td>&gt;10</td>
<td>0</td>
</tr>
</tbody>
</table>
### Table 3. Sociodemographic characteristics of informal caregivers.

<table>
<thead>
<tr>
<th>ID</th>
<th>Sex</th>
<th>Age (years)</th>
<th>Employment status</th>
<th>Relationship with the person with dementia</th>
<th>Living situation of the person with dementia</th>
<th>Years of caring experience</th>
<th>The person with dementia has passed away</th>
<th>Focus group attended⁵</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Female</td>
<td>58</td>
<td>Housewife or retired</td>
<td>Spouse</td>
<td>Own residence</td>
<td>3-5</td>
<td>No</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
<td>55</td>
<td>Housewife or retired</td>
<td>Daughter</td>
<td>Own residence</td>
<td>3-5</td>
<td>No</td>
<td>1, 3, and 4</td>
</tr>
<tr>
<td>3</td>
<td>Female</td>
<td>59</td>
<td>Housewife or retired</td>
<td>Spouse</td>
<td>Own residence</td>
<td>3-5</td>
<td>No</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Male</td>
<td>67</td>
<td>Employed</td>
<td>Son</td>
<td>Own residence</td>
<td>3-5</td>
<td>No</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>Female</td>
<td>58</td>
<td>Housewife or retired</td>
<td>Spouse</td>
<td>Caregiver's residence</td>
<td>1-2</td>
<td>No</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>Male</td>
<td>57</td>
<td>Employed</td>
<td>Son</td>
<td>Own residence</td>
<td>3-5</td>
<td>No</td>
<td>1 and 4</td>
</tr>
<tr>
<td>7</td>
<td>Male</td>
<td>74</td>
<td>Employed</td>
<td>Son</td>
<td>N/A⁻¹</td>
<td>&gt;10</td>
<td>Yes</td>
<td>1 and 2</td>
</tr>
<tr>
<td>8</td>
<td>Female</td>
<td>55</td>
<td>Employed</td>
<td>Daughter</td>
<td>Own residence</td>
<td>3-5</td>
<td>No</td>
<td>2 and 4</td>
</tr>
<tr>
<td>9</td>
<td>Female</td>
<td>75</td>
<td>Housewife or retired</td>
<td>Daughter</td>
<td>N/A</td>
<td>&gt;10</td>
<td>Yes</td>
<td>1, 2, 4, and 5</td>
</tr>
<tr>
<td>10</td>
<td>Female</td>
<td>76</td>
<td>Housewife or retired</td>
<td>Spouse</td>
<td>Own residence</td>
<td>3-5</td>
<td>No</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>Female</td>
<td>82</td>
<td>Housewife or retired</td>
<td>Spouse</td>
<td>Caregiver's residence</td>
<td>1-2</td>
<td>No</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>Female</td>
<td>55</td>
<td>Employed</td>
<td>Daughter</td>
<td>Own residence</td>
<td>1-2</td>
<td>No</td>
<td>3 and 4</td>
</tr>
<tr>
<td>13</td>
<td>Female</td>
<td>81</td>
<td>Housewife or retired</td>
<td>Spouse</td>
<td>Caregiver's residence</td>
<td>3-5</td>
<td>No</td>
<td>2 and 4</td>
</tr>
</tbody>
</table>

¹Number of the focus group attended.

⁻¹N/A: not applicable.

**FG With Formal Caregivers**

Formal caregivers are professionals who are trained, hired, and paid to provide care to a person living with dementia. In our study, all formal caregivers (n=6) actively participated in the digital discussions. All participants were female. Their mean age was 44 (range 28-59) years. All participants had professional experience in taking care of a person living with dementia. Of the 6 participants, 3 (50%) reported having >10 years of experience in dementia care. In addition to the professional caregiving experience, 3 (50%) of the 6 participants also reported taking care, or having taken care in the past, of a relative affected by dementia (Table 2). The main findings of the FG are summarized in the following paragraphs.

Participants agreed that an intervention aiming to support and improve the knowledge of informal caregivers of people with dementia was much needed. A caregiver compared information learning to a safeguard not only for the carer but also for the care recipient:

> I hope that this program will spread because information protects all of us: the carer, and especially the person who is cared for. [ID 6]

iSupport was generally appreciated and acknowledged by participants as a useful tool. The contents were found appropriate and sufficiently comprehensive. The difficulty regarding accepting the disease and the changes in the relationship with the care recipient were found to be the main challenges and contents to cover in the program:

> Relates find it extremely difficult to accept the disease and the change...I believe a very strong support is needed...also at a social level because the disease is often associated with shame. [ID 1]

Similarly, a participant also suggested adding to the program specific resources for social and psychological support:

> You could mention [the existence or the opportunity for family members] to benefit from psychological support because they need it, always. [ID 2]

This quote underscored the recognition by formal caregivers of the potential emotional and psychological strain on family members as they witness the progression of the disease of their care recipients.

In light of the participants’ perspectives, an important feature to add to the original iSupport format was the inclusion of a platform for caregivers to engage with each other and that facilitated the caregivers’ interactions with each other (this adaptation was also needed to differentiate iSupport from another repository of information or digital available resources on dementia):

> There are a billion guides on dementia...I think people need to interact. [ID 4]
Regarding case scenarios, the caregivers generally found that the examples were appropriate and consistent with their experiences. However, the answer options often did not reflect the variety of, and differences in, caregiving situations and experiences, including the age of the person affected by dementia, the severity of dementia, the living situation, or the type of dementia (eg, Alzheimer disease and frontotemporal dementia). A participant suggested adding general guidelines to the examples to include more answers:

*If the examples aim to increase knowledge, they should give general indications that can apply to different caring situations.* [ID 6]

### FG With Informal Caregivers

Of the 20 informal caregivers who contacted us, 13 (65%) joined the FGs. Reasons for nonparticipation were lack of time and geographic distance. Most of the caregivers (10/13, 77%) were female; nearly half were spouses (6/13, 46%) of the persons living with dementia, and more than half were children (7/13, 54%; daughter: n=4, 57%; son: n=3, 43%) of the persons living with dementia. Their age ranged from 55 to 82 years. Most of the participants (10/13, 77%) reported a caregiving experience of at least 3 years, and most of them (8/13, 62%) cared for a relative who lived at their own residence. Of the 13 participants, 2 (15%) reported that the person they cared for had passed away.

The number of caregivers attending each FG ranged from 2 to 7: of the 13 caregivers, 6 (46%) attended FG 1 on June 14, 2021; a total of 7 (54%) attended FG 2 on July 12, 2021; a total of 2 (15%) attended FG 3 on July 15, 2021; a total of 6 (46%) attended FG 4 on August 18, 2021; and 2 (15%) attended FG 5 on August 24, 2021. Of the 13 participants, 7 (54%) attended >1 FG (Table 3). The main findings of the 5 FGs are summarized in the following paragraphs.

Participants generally believed that iSupport holds the promise to be useful, to increase dementia knowledge, and provide information about available services and support measures for people living with dementia and their families:

*The idea is brilliant because everything can be useful...In my opinion, the most interesting thing is the overview of what is locally available to support caregivers.* [ID 9; spouse]

The need for guidance and orientation to services was felt owing to a perceived lack of support and direction, likely stemming from the uncertainty and confusion that frequently followed the diagnosis. This feeling of bewilderment was echoed by a participant:

*It's confusing outside, you don’t know where to go, whom to turn to...there are no guidelines, no support.* [ID 2; daughter]

Regarding the contents, participants reported familiarity with most of the case scenarios.

A participant commented as follows on a scenario (module 3, chapter 3) involving a person affected by dementia who cannot find the house keys and does not want the carer to leave him alone at home:

*It happened to me many times, not always with the keys though.* [ID 12; daughter]

However, despite the familiarity reported and the need to obtain information and increase knowledge to cope with difficult situations, the original exercise format was seen as a limitation by some of the participants. A participant reported feeling diminished when choosing between wrong and right answers:

*It’s almost guilt-inducing...There is the best solution and if you guess wrong you are doing your role wrong.* [ID 8; daughter]

In addition, some of the answer options were considered to be so wrong as to be offensive to the carer; for example, in module 5, chapter 9, a case scenario described a situation where the person affected by dementia (Matteo) makes sexual remarks toward a domestic worker, and the user is asked what they would do in this situation. A participant commented on the option “shout at Matteo and shame him for his conduct” as inconceivable:

*Shout?!...We do know what we’re doing!* [ID 9; daughter]

Similar to what formal caregivers reported about case scenarios, participants also highlighted the risk of generalizing solutions that may not be appropriate for all caregiving situations:

*It should be clear that each user has to transpose his or her situation by taking cues from the scenario, but unfortunately it isn’t black and white.* [ID 13; spouse]

Finally, the informal caregivers too suggested adding interactive features to the digital version of iSupport to minimize the risk of the caregivers isolating themselves; for instance, a participant commented as follows:

*For me, the biggest utility is in connecting people...there should be people behind the app.* [ID 5; spouse]

### Discussion

#### Overview

This study described in detail the main steps taken to culturally adapt the WHO iSupport program for informal caregivers of people living with dementia in Switzerland. Our results demonstrate the complexity as well as the necessity of adapting an evidence-based complex intervention to a specific cultural context and population. We collected feedback and implemented changes, in accordance with the WHO authors of the program, to the original iSupport version in the areas of the language, resources, contents, graphics, and features used in the program.

In the following paragraphs, we summarize and comment on the main lessons learned.

#### Valuing Experiential Knowledge

One of the main messages we took away during the adaptation process was the importance placed by informal caregivers on being recognized for their role and expertise. This finding is consistent with the findings of other studies, including the works reporting on iSupport adaptation processes in other countries [31,34,43-45]. Our participants suggested that the learning
approach used in the original iSupport program was too scholastic and recommended the removal of expressions that likely resulted from a top-down approach to content and compilation. Referring to case scenarios, some of the informal caregivers (5/13, 38%) felt that the simplicity of certain answer options was offensive. Informal caregivers claimed to be recognized because of their lived experience as experts in the field who could contribute to not only locally adapting iSupport but also integrating and shaping it. This echoes the inclusive procedures used to develop iSupport in the first place [21] and the work done for the iSupport adaptation process in Portugal and the United Kingdom [34,43]. Informal caregivers can spend on average 170 hours a month providing care to a loved one affected by dementia [46]. In our study, more than half of the caregivers (10/13, 77%) reported a caregiver experience of at least 3 years and up to 10 years. Although one may argue that caregivers acquire and improve their learning by doing, it is undeniable that they can become experts in caring; surely, they provide a unique perspective of the person with dementia and their own needs. However, besides the years of personal experience, caregivers’ knowledge of dementia and caring may also depend on other factors and may be influenced by their educational level and sociocultural background. Similar to any complex health intervention [47], it is important to ensure that the final version of iSupport is adapted to the real user’s experience and preexisting abilities. An early, timely, and active involvement of caregivers is needed [48,49]. The adoption of a language register and skills training techniques that promote preexisting abilities, rather than replace them, may enhance the acceptance and use of the intervention.

Enhancing Social Contacts

According to participants, iSupport could benefit from the inclusion of interactive features (eg, chat and forum) that allow the user to communicate with other caregivers and share experiences and problem-solving strategies. This finding is consistent with a recent study [50] that found that peer support can be complementary to professional support and beneficial in reducing social isolation, as well as in connecting patients and caregivers to others with similar issues. Similarly, Greenwood et al [51] found that, besides providing psychosocial support, peer support interactions for caregivers of people with dementia can offer practical information and guidance in managing difficult situations and gaining new perspectives on their caring role.

The adoption of peer support programs for informal caregivers of people with chronic diseases and disabilities is well established in the literature [52]. A recent scoping review [53] found that peer support was often part of multicomponent interventions that also addressed information sharing, skills development, personal coping skills, and self-management. Despite the difficulty in identifying what component may or may not be beneficial for the carers, the authors concluded that peer support, particularly if delivered digitally, could represent a cost-effective medium and opportunity to meet caregivers’ needs and preferences.

Importantly, digital meets among peers seem more promising, usable, and potentially effective for caregivers when embedded in digital interventions [10] such as iSupport.

Facilitating Access to, and Navigation of Local Services

Another suggested feature to implement in the program was the inclusion of contacts of local resources for dementia, such as health care services and facilities, charities, or other relevant organizations. Consistent with what our participants reported, informal caregivers often experience a lack of information and support, especially at the beginning of the caregiver journey, when it is best to establish fruitful contacts and interactions with local health and social care services and offers in general [1]. According to the latest World Alzheimer Report [3], <50% of informal caregivers are advised to contact the local Alzheimer association or receive postdiagnostic support information. The navigation of the services and various offers for both people living with dementia and informal caregivers is taxing, often ineffective, and can be frustrating. The lack of information about existing services and support is associated with caregiver burden and distress [54]. A recent review on the needs of family caregivers revealed that information provided on available support services and measures was one of the main needs reported by caregivers after their loved one was diagnosed with dementia [55]. Caregivers may seek support autonomously, mainly digitally. However, the variety of information and sources available on the internet about dementia may contribute to creating feelings of bewilderment and difficulties in finding relevant and reliable information [56]. Hence, digital interventions that also include contacts with external and local resources may help users to access and navigate the health care system and find the most appropriate service or information for their situation.

Limitations

We acknowledge that our study has limitations. First, we included only a few participants for each phase of the adaptation process. Because of their pressing needs and duties, informal caregivers are a challenging population to reach and involve in research [57]. However, the number of caregivers and experts that we included in our study was adequate for the qualitative methods used and is higher than the minimum recommended by the WHO guidelines to adapt iSupport to local contexts [22]. In addition, we set up a CAB that included both stakeholders and caregivers who worked continually and with great dedication through the adaptation process of iSupport. Second, the discrepancy in FG size between formal and informal caregivers and the attendance of informal caregivers in >1 FG may have contributed to reaching data saturation, but this may have reduced social desirability bias, thanks to both the progressive cementing of positive small group dynamics among participants and the variety of the contents discussed. Third, our study was conducted in Switzerland, a high-income country, equipped with a National Dementia Strategy that aims to improve the quality of life of people affected by dementia and to promote awareness and education on dementia [58]. Therefore, the feedback and experiences that we collected may not be easily generalized to all contexts. However, the adaptation strategies and phases described in our study may be useful for
all countries interested in adapting digital interventions for caregivers of people with dementia, not only iSupport. Our findings suggest that digital interventions benefit from a community-based participatory approach and the involvement of caregivers to ensure that the final program meets the needs and preferences of users [17].

Future Research
The recommendations and feedback that we collected during this study allowed us to adapt the original contents of the iSupport program to the Swiss context and to inform the development of the iSupport desktop version, mobile app, and printed manual. Following the Medical Research Council guidelines for the development of complex health interventions [59], we will proceed to assess the usability and feasibility of iSupport before its implementation. Evidence not only on the effectiveness but also on the ease of implementation and scalability of caregivers’ interventions is still rare in our country. We are determined to design and conduct good-quality studies to address these gaps and to promptly disseminate our findings and experience widely through peer-reviewed publications, the WHO knowledge exchange platform [60], and the global WHO iSupport network coordinated by the Brain Health Unit at the WHO.

Finally, the iSupport original program was developed by the WHO based on evidence related to carer training and support interventions and in collaboration with experts and caregivers [21]. Therefore, the program can be adapted to the extent that it maintains the original aims and structure [22]. During the study, we collected recommendations and feedback that would have required a consistent change in terms of resources and digital infrastructure to be implemented. These included, for instance, contents based on the type of dementia and stage of the disease, a comprehensive map of all digital and local resources available, and consultation from professionals as well as legal and financial assistance. Therefore, further development of iSupport could focus on supporting specific groups of caregivers, such as young carers or caregivers of people with rare dementia, and on providing personalized support tailored to the stage of the caregiver journey and the care needs of the care recipient.

Conclusions
Despite the recognized importance of culturally adapting interventions to implement them in real-world settings, the evidence on how to conduct this process is still limited. Our study enriches this landscape by underscoring that an active engagement of the final users and stakeholders allows to adapt an intervention to their culture, values, and needs. In addition, this study provides examples of concrete strategies and methods to involve community members and stakeholders across different phases of the intervention. Indeed, despite the emerging importance of coconstructing research together with people as collaborators, rather than as simply subjects of traditional research, there is limited evidence regarding the modalities of this practice.

Our experience confirms that the adoption of a CBPR approach is necessary to identify and address criticisms and potential barriers to the use and acceptance of a digital educational intervention before its implementation. In conclusion, we envision this study as a potential driver for enhancing a more robust dialogue between researchers and communities. We firmly believe that CBPR represents a transformative research opportunity where the needs of academics and community members can be met and where both groups can find opportunity for mutual knowledge exchange and growth.

Acknowledgments
The authors thank all informal and formal caregivers of people with dementia for their time and precious contribution to this work. The authors also wish to acknowledge the support provided by the members of the community advisory board, the funders, and collaborators for the realization of this project.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Community advisory board agreement.
[PDF File (Adobe PDF File), 187 KB - formative_v8i1e46941_app1.pdf ]

Multimedia Appendix 2
Informed consent.
[PDF File (Adobe PDF File), 246 KB - formative_v8i1e46941_app2.pdf ]

Multimedia Appendix 3
World Health Organization adaptation spreadsheet.
References


Abbreviations

CAB: community advisory board
CBPR: community-based participatory research
FG: focus group
REDCap: Research Electronic Data Capture

https://formative.jmir.org/2024/1/e46941

PMID: 38705532
DOI: 10.2196/24186
Culturally Adapting the World Health Organization Digital Intervention for Family Caregivers of People With Dementia (iSupport): Community-Based Participatory Approach

Messina A, Amati R, Annoni AM, Bano B, Albanese E, Fiordelli M

Please cite as:
Messina A, Amati R, Annoni AM, Bano B, Albanese E, Fiordelli M
Culturally Adapting the World Health Organization Digital Intervention for Family Caregivers of People With Dementia (iSupport): Community-Based Participatory Approach
JMIR Form Res 2024;8:e46941
URL: https://formative.jmir.org/2024/1/e46941
doi:10.2196/46941
PMID:38265857

©Anna Messina, Rebecca Amati, Anna Maria Annoni, Beatrice Bano, Emiliano Albanese, Maddalena Fiordelli. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
A Bluetooth-Enabled Device for Real-Time Detection of Sitting, Standing, and Walking: Cross-Sectional Validation Study

Reza Daryabeygi-Khotbehsara1, MSc, PhD; Jonathan C Rawstorn1, PhD; David W Dunstan2, PhD; Sheikh Mohammed Shariful Islam1, MD, PhD; Mohamed Abdelrazek3, PhD; Abbas Z Kouzani4, PhD; Poojith Thummala1, MSc; Jenna McVicar1, MSc, PhD; Ralph Maddison1, PhD

1Institute for Physical Activity and Nutrition, School of Exercise and Nutrition Sciences, Deakin University, Melbourne Burwood, Australia
2Baker-Deakin Department of Lifestyle and Diabetes, Melbourne Burwood, Australia
3School of Information Technology, Deakin University, Melbourne Burwood, Australia
4School of Engineering, Deakin University, Geelong, Australia

Corresponding Author:
Reza Daryabeygi-Khotbehsara, MSc, PhD
Institute for Physical Activity and Nutrition, School of Exercise and Nutrition Sciences
Deakin University
221 Burwood Hw
Melbourne Burwood, 3125
Australia
Phone: 61 3 924 45936
Email: reza.d@deakin.edu.au

Abstract

Background: This study assesses the accuracy of a Bluetooth-enabled prototype activity tracker called the Sedentary behaviOR Detector (SORD) device in identifying sedentary, standing, and walking behaviors in a group of adult participants.

Objective: The primary objective of this study was to determine the criterion and convergent validity of SORD against direct observation and activPAL.

Methods: A total of 15 healthy adults wore SORD and activPAL devices on their thighs while engaging in activities (lying, reclining, sitting, standing, and walking). Direct observation was facilitated with cameras. Algorithms were developed using the Python programming language. The Bland-Altman method was used to assess the level of agreement.

Results: Overall, 1 model generated a low level of bias and high precision for SORD. In this model, accuracy, sensitivity, and specificity were all above 0.95 for detecting sitting, reclining, standing, and walking. Bland-Altman results showed that mean biases between SORD and direct observation were 0.3% for sitting and reclining (limits of agreement [LoA]=–0.3% to 0.9%), 1.19% for standing (LoA=–1.5% to 3.42%), and –4.71% for walking (LoA=–9.26% to –0.16%). The mean biases between SORD and activPAL were –3.45% for sitting and reclining (LoA=–11.59% to 4.68%), 7.45% for standing (LoA=–5.04% to 19.95%), and –5.40% for walking (LoA=–11.44% to 0.64%).

Conclusions: Results suggest that SORD is a valid device for detecting sitting, standing, and walking, which was demonstrated by excellent accuracy compared to direct observation. SORD offers promise for future inclusion in theory-based, real-time, and adaptive interventions to encourage physical activity and reduce sedentary behavior.

(JMIR Form Res 2024;8:e47157) doi:10.2196/47157

KEYWORDS
activity tracker; algorithms; deep neural network; machine learning; real-time data; Sedentary behaviOR Detector; sedentary behavior; SORD; standing; validation; walking; wearables

Introduction

Sedentary behavior (SB) is defined as “any waking behavior characterized by an energy expenditure of less than 1.5 metabolic equivalents while in a sitting, reclining, or lying posture” [1,2]. SB is an independent risk factor for many noncommunicable diseases, with the risk being most pronounced in those who are also physically inactive (ie, not meeting physical activity [PA] guidelines) [3-5]. Reducing SB for all
people, including those who are physically active, can assist in producing health benefits [6]. Interrupting SB with standing or light or moderate intensity PA can improve chronic risk factors including glucose homeostasis, insulin sensitivity, blood lipid concentrations, and diastolic blood pressure [7-10]. Recent World Health Organization guidelines on PA and SB explicitly state the importance of reducing sedentary time in addition to promoting PA for adults and older adults, including those with chronic conditions [11]. This has subsequently led to the development of interventions targeting SB reduction, although interventions to date have been compromised by the lack of a tool that can capture SB accurately and in real time. Accurate measurement of sitting, standing, and walking in real time will enable the design of interventions that can adapt to changes in the activity state and can be delivered at times when an individual is most responsive to the intervention, therefore maximizing the potential opportunity for reducing SB and increasing PA [12].

To date, the majority of interventions to reduce SB and promote PA have relied on subjective measurement of these behaviors, which are subject to self-report bias [13,14] and may underestimate daily sitting time by up to 2 hours compared with objective measurement [15]. Few activity trackers, including research-grade (eg, activPAL) and commercial (eg, GENeActive and Fitbit One), measure sedentary time with reasonable precision [16-22], but they are not optimal for SB change interventions [23]. The 2 main issues involve technical difficulty in using support software for real-time interventions and concerns about device accuracy in distinguishing postural states (sitting, standing, etc) [24,25]. Most activity trackers use similar technologies, including accelerometers, magnetometers, and gyroscopes, to detect posture and activity [26]. However, the placement of devices on the body can considerably influence accuracy [27]. Commercial wrist-worn devices such as the Garmin Vivofit are unable to detect sit-to-stand transition [23,28]. Other thigh-worn devices, such as activPAL and SitFIT, are capable of detecting sitting and standing due to their horizontal placement [23]. In terms of behavioral intervention, activPAL does not offer any real-time prompts or feedback to participants [29]. The SitFit device provides real-time feedback to the user, and its accuracy, although acceptable, was lower when compared to the activPAL, which is considered the preferred device for research purposes [30]. However, SitFit is pocket-worn, which limits its use for those not wearing suitable clothing (eg, trousers) or garments without pockets (eg, dresses) [30]. More importantly, SitFit does not distinguish standing from walking [31] and therefore cannot be used to assess standing as a unique outcome both for real-time and adaptive interventions. It should be noted that these devices (SitFit and Fitbit One) are no longer available on the market and were included in our discussion to provide historical context and illustrate the evolution of activity-tracking technology. Evidence on the positive impact that standing may have on health outcomes in different population groups is emerging from short-term and small-scale studies [32,33], although real-time assessment and behavior change interventions are missing. This, in turn, suggests a need for a platform to momentarily evaluate both sedentary and standing outcomes to study their exclusive health effects and intervene accordingly.

In summary, despite the presence of activity tracker devices, few have included evidence- and theory-based interventions or strategies to promote PA and reduce SB (eg, self-monitoring and goal setting), and the use of some other devices is restricted due to a lack of real-time assessment of outcomes (eg, standing). In response, we designed and developed a new wearable platform called “Sedentary behaviOR Detector” (SORD), which collects real-time sedentary data, including lying, reclining, sitting, and standing, as well as walking activity time. Therefore, this study aimed to assess the validity of the SORD device in detecting sedentary and walking activities among adult participants.

Methods

Overview

A cross-sectional, laboratory-based study was conducted to assess the criterion validity (SORD vs direct observation) and convergent validity (SORD vs activPAL). Adults were recruited to take part in this laboratory-based study through print and email advertisements at a university campus. Adults aged 18 years or older, without gait abnormalities, able to walk on a treadmill easily, with no skin sensitivity to plasters or tapes, and able to communicate in English were included.

Upon arrival, participants completed a demographic questionnaire including age, sex, ethnicity, job status, marital status, education, and the Physical Activity Readiness Questionnaire [34] for safe exercise. Anthropometric measures, including height to the nearest 0.1 cm and weight to the nearest 0.1 kg, were taken using a stadiometer (Seca 213) and Tanita scale (Tanita Innerscan 50), respectively.

Participants were given a printed activity protocol to help familiarize them with the required activities and the order in which they were to be performed. Textbox 1 presents a range of different states of activities included in the study protocol to mimic typical postures that may be encountered during everyday life.

Hypoallergenic retention dressing tape (Hypafix) was used to attach the SORD and activPAL devices on the midline of the right thigh. Participants were then instructed to engage in a combination of activities in the order of sitting, reclining, sitting, standing, walking, sitting, lying, and walking on a treadmill. Each activity variation lasted for a minimum of 2 minutes and a maximum of 3 minutes and 30 seconds, except walking, which involved participants walking at their regular walking pace along a 10-m-long path. Participants had 2 minutes of optional resting to break up the activities if needed. Ground truth, or the true time spent on each of the activities, was measured by a researcher with the help of a video camera for direct observation.
Textbox 1. Details of the Sedentary behaviOR Detector phase 1 activities.

<table>
<thead>
<tr>
<th>Lying</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Face up, on the right shoulder, face down, or on the left shoulder</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reclining</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Normal (135 slope chair), left leg over right, or right leg over left</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sitting</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Upright, ankle-on-knee (left-right and right-left), right foot move, left foot move, both feet move, elbows on legs, or sitting with outstretched legs</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standing</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Stand normal, casual standing (more weight on the right foot), casual standing (more weight on the left foot), right shoulder on the wall, or left shoulder on the wall</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Walking</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Normal on level, on treadmill at 4 km/h, or on treadmill at 6 km/h</td>
<td></td>
</tr>
</tbody>
</table>

Sedentary behaviOR Detector

Overview

The SORD is a wearable electronic device (Figure 1A) that collects and provides real-time data associated with sitting, reclining, lying, and PA. Data provided by the device can be used to separate sitting versus standing versus ambulation. To separate sitting time from lying time, 2 same devices will be attached to 2 different locations of the body.

The SORD device includes a number of internal components (Figure 1B): a low-power processor and transceiver, inertial measurement unit, voltage regulator, battery charger, battery, antenna, micro-USB connector, LEDs, motherboard, and an enclosure. These components have been described below.

Figure 1. (A) Sedentary behaviOR Detector (SORD). (B) Internal components of the SORD device. The SORD is a small device with the following dimensions: 0.9 mm (height), 37 mm (width), and 68 mm (length). It is also lightweight, with a weight of 23.5 g. The device can operate for about 45 hours on a single charge. The SORD device measures 3-axis orientation using the accelerometer that gives acceleration signals for 3 axes, the gyroscope that provides rotation along 3 axes, and the magnetometer that gives motion in the magnetic field in 3 axes. It hosts an embedded C firmware that continuously reads from the sensors, records their data at 25-Hz frequency, preprocesses the data, and transmits the data wirelessly. No initialization is required for the SORD device, as the data are captured and transmitted through the 2.4-GHz Bluetooth Low Energy 5.0 transceiver in real time.

Processor and Transceiver

The ATSAMB11-ZR210CA is used that includes a low-power ARM Cortex M0 32-bit processor, 128 KB of RAM, 128 KB of stacked flash memory, a 2.4 GHz Bluetooth Low Energy 5.0 transceiver and modem, a power management unit, a ceramic high-gain antenna, and a printed circuit board with a small footprint.

Inertial Measurement Unit

The BNO055 is used that includes a single-chip integrated circuit incorporating an intelligent inertial measurement unit with a triaxial 14-bit accelerometer, a triaxial 14-bit gyroscope, a triaxial geomagnetic sensor, an I2C communication interface, and an ARM Cortex M0+ 32-bit processor executing a sensors data fusion algorithm.

Voltage Regulator

The XC9264B755MR-G is used which includes a synchronous step-down DC/DC voltage regulator. It operates within the voltage range of 3-18 V and provides a 500 mA output current. It has a selectable switching frequency of 500 kHz, 1.2 MHz, or 2.2 MHz. It also features overcurrent protection as well as thermal shutdown.
**Battery Charger**
The BQ25101YFPR is used which includes a linear Li-Ion and Li-Pol battery charger with a very small footprint. It has a single power output that charges a battery in 3 steps: conditioning, constant current, and constant voltage. The junction temperature of the device is monitored to control the charge current.

**Micro-USB Connector**
A micro-USB connector is used for programming the processor and also for establishing serial communications as well as charging the onboard battery.

**LEDs**
A total of 2 multicolor LEDs are used to illuminate different functional states of the device to the user.

**Battery**
A 3.7-V, 700-mAh, 303759 Lithium Polymer rechargeable battery is used. Its height, width, and length are 3 mm, 37 mm, and 59 mm, respectively, and its weight is 14 g.

**Antenna**
A Freedom 2.4-GHz flex circuit PCB antenna is used.

**Motherboard**
A printed circuit motherboard is designed and fabricated to host all the electronic components of the SORD device.

**Enclosure**
A small enclosure for the SORD device is designed and 3D printed. It hosts all the components of the device.

**ActivPAL**
ActivPAL is a thigh-worn triaxial accelerometer that classifies an individual’s activity into periods of time spent sedentary (lying or sitting), standing, and walking, as well as the number of steps and stepping speed [29,35]. ActivPAL devices were initialized before the data collection and date-time stamped 1-second epoch files were used for comparative analysis.

**Direct Observation**
True time spent engaging in activities was logged by a trained researcher (RDK). This was assisted by a video camera positioned in the room and checked by another researcher (JM). If there was any discrepancy, RDK and JM reviewed the camera data together to achieve consensus. No formal intra- or interrater reliability was conducted.

**Data Handling and Analysis**
SORD data were transmitted to a computer through Bluetooth Low Energy. A program was developed in MATLAB (MathWorks) and run on a Microsoft Windows (Microsoft Corp.)–based computer to receive data from the SORD devices in real time and store it into a Microsoft Excel (Microsoft Corp.) file. The program starts by initializing relevant variables and a communications port, creates a file name based on the current date and time, continuously receives data from the SORD devices, and stores the incoming data in the Microsoft Excel file in real time. Each data packet received from the SORD devices includes values obtained from the onboard sensors at the current time. For a data packet, the following information is then stored in the file in real time: date, time, angle, accelX, accelY, accelZ, gyroX, gyroY, gyroZ, magX, magY, magZ, and battery voltage. To avoid potential Bluetooth transmission package loss, this study used the time-stamp data from the SORD device instead of the computer receiver (ie, the sending time stamp rather than the receiving time stamp). Thus, we had computer receiver and accelerometer data, along with their timestamps. Based on the real sampling rate, the computer calculated the time stamp difference between each data point (ΔT). In this research, the number of missing data points was defined by missing = ΔT/(1/25Hz) – 1. The values of these data points were filled by the average of the 2 data points before and after the missing data points (eg, Vi[missing] = [Vi – 1 + Vi + 1]/2). Before sending data to the server for inference, the phone app waits until all required data have been received (processing buffer length). Using the VANE (standard) classification algorithms, activPAL data were processed and collected using proprietary software (activPAL Professional Research Edition, PAL Technologies). The software-generated event file was used. This file contains a chronological list of all episodes of sedentary, standing, and stepping (ie, walking) activities recorded at 1-second intervals. The frequency of the recorded signals from SORD was subsequently reduced to 1 Hz (ie, 1-second epochs) for comparative analysis. This reduction in frequency simplifies data processing and facilitates direct comparison with activPAL, which was also sampled at 1 Hz. Furthermore, outliers or irregular data points were identified and removed. Once individual data sets were cleaned, they were combined for subsequent comparative analysis. The combining process involved aligning the data sets temporally so that corresponding data points from both devices were synchronized for direct comparison.

Due to multiple limitations, we did not use the available open-source activity recognition algorithms. These limitations include (1) inconsistency in the data format and ranges, (2) differences in the frequency of raw data assumed by these algorithms compared to SORD (which is 28 measurements per second), and (3) the variations of activities considered by these algorithms were not exactly the same as the ones we wanted to address in this research at this stage and in the future. Thus, we developed the data engineering and activity recognition models. Although ensemble learning techniques outperform deep learning, they demand higher computation resources and have longer processing latency [36,37]. Therefore, for practical reasons and real-world applications of SORD, this study used deep neural network models—a combination of convolutional neural network and recurrent neural network—to develop algorithms. Deep neural network can learn features automatically from the raw data, therefore performing better than statistical and basic machine learning methods, and they are suitable for recognizing complex activities [38]. A data scientist developed deep learning algorithms to classify activity type and postural states from preprocessed motion sensor data using the Python programming language [39]. First, machine learning classifiers were developed, trained, and tested for the SORD device. A dynamic sliding window approach was used for machine learning [40], where each window was related
to a particular activity and multiple variables were examined within each window to identify patterns. When a particular activity was detected in the sensor readings, features were extracted to classify activities between the previous one and the current one (further details are provided below). Then, criterion validity (against direct observation) and convergent validity (against activPAL micro) were evaluated. Using Python, the Bland-Altman method was used to assess the level of agreement between SORD and each reference measure (criterion agreement=directly observed time and convergent agreement=activPAL). Mean difference represents the systematic bias, and the limits of agreement (LoA) show the range of agreement between SORD and reference methods, where a positive value indicates underestimation and a negative value indicates overestimation by SORD. For all activity states, we predefined the acceptable LoA between ±10%.

**Classification Algorithms**

A single data set included SORD, activPAL, and direct observation data for 1 participant. Deep learning was used to randomly select 6 data sets for training, 1 for validation, and 7 for testing. In the training set, similar patterns were identified for the previous 35 data points to specify an activity. Confusion matrices were used to visualize the model’s performance. In a confusion matrix, each row represents the instances in the predicted activity, and each column represents the instances in the actual activity.

**Ethical Considerations**

Ethics approval was granted by the Deakin University Human Research Ethics Committee’s Human Ethics Advisory Group (HEAG-H 109_2019). All participants provided written informed consent. All research data were anonymized before cleaning and analysis. Participants were remunerated with an Aus $20 (US $14) gift voucher.

**Results**

**Overview**

In total, 15 adults (12 female adults) aged between 20 and 62 years completed the experimental study. Table 1 presents the demographic characteristics of the participants.
Table 1. Demographic information of study participants.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>Median (SD)</td>
<td>35.2 (11.6)</td>
</tr>
<tr>
<td>Range</td>
<td>20-62</td>
</tr>
<tr>
<td><strong>Weight (kg)</strong></td>
<td></td>
</tr>
<tr>
<td>Median (SD)</td>
<td>70.4 (10.5)</td>
</tr>
<tr>
<td>Range</td>
<td>55.2-84.8</td>
</tr>
<tr>
<td><strong>Height (cm)</strong></td>
<td></td>
</tr>
<tr>
<td>Median (SD)</td>
<td>168.1 (9.6)</td>
</tr>
<tr>
<td>Range</td>
<td>147.0-186.5</td>
</tr>
<tr>
<td><strong>BMI (kg/m²)</strong></td>
<td></td>
</tr>
<tr>
<td>Median (SD)</td>
<td>24.9 (3.0)</td>
</tr>
<tr>
<td>Range</td>
<td>20.1-29.4</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>12 (80)</td>
</tr>
<tr>
<td>Male</td>
<td>3 (20)</td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Australian</td>
<td>4 (27)</td>
</tr>
<tr>
<td>European</td>
<td>6 (40)</td>
</tr>
<tr>
<td>Asian</td>
<td>4 (27)</td>
</tr>
<tr>
<td>South American</td>
<td>1 (7)</td>
</tr>
<tr>
<td><strong>Education level, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Degree higher than bachelor’s (bachelor’s with honors, masters, or PhD)</td>
<td>7 (47)</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>5 (33)</td>
</tr>
<tr>
<td>Technical and further education or university course below a bachelor’s degree</td>
<td>2 (13)</td>
</tr>
<tr>
<td>Other school qualifications (eg, overseas school, Cambridge examination, or A level)</td>
<td>1 (7)</td>
</tr>
<tr>
<td><strong>Job status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Full-time salary or wage earner</td>
<td>6 (40)</td>
</tr>
<tr>
<td>Part-time salary or wage earner</td>
<td>2 (13)</td>
</tr>
<tr>
<td>Student</td>
<td>7 (47)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Married or living with a partner</td>
<td>8 (53)</td>
</tr>
<tr>
<td>Single or never married</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Separated, divorced, or widowed</td>
<td>6 (40)</td>
</tr>
</tbody>
</table>

Deep Learning Results

A total of 4 models were presented for SORD. Model 1 classified 3 activities, including sedentary (lying, reclining, or sitting), standing, and walking separately. As illustrated in Figure 2, model accuracy, sensitivity, and specificity for detecting sedentary time were 0.92, 0.99, and 0.87; for standing, they were 0.95, 1.00, and 0.91; and for walking, they were 0.96, 0.92, and 1.00, respectively.

Model 2 included 4 activities: sitting, reclining, standing, and walking; lying was excluded (ie, lying moments observed by video camera were omitted from the data set). As illustrated in Figure 3, model accuracy, sensitivity, and specificity for detecting sitting and reclining were 1.00, 1.00, and 1.00; for standing, they were 0.99, 0.99, and 1.00; and for walking, they were 0.98, 1.00, and 0.95, respectively.

Model 3 included 3 activities: sitting, standing, and walking; reclining and lying were excluded. Respectively, model accuracy, sensitivity, and specificity for detecting sitting were...
0.97, 1.00, and 0.94; for detecting standing, they were 0.95, 0.91, and 1.00; and for walking, they were 0.98, 1.00, and 0.97 (Multimedia Appendix 1).

Model 4 included all 5 activities: lying, sitting, reclining, standing, and walking. Respectively, model accuracy, sensitivity, and specificity for detecting lying were 0.70, 0.54, and 1.00; for sitting and reclining, they were 0.85, 1.00, and 0.75; for standing, they were 0.75, 0.63, and 0.93; and for walking, they were 0.99, 1.00, and 0.98 (Multimedia Appendix 2).

Figure 2. Confusion matrix for model 1 classification algorithms. Sedentary (lying, sitting, and reclining), standing, and walking were included in the model.

![Confusion Matrix](image1)

Figure 3. Confusion matrix for model 2 classification algorithms. “Sitting and reclining,” standing, and walking were included in the model.

![Confusion Matrix](image2)
Agreement

For models 1 and 2, results of the Bland-Altman analysis comparing second-by-second data on sedentary, standing, and walking time between direct observation versus SORD and activPAL versus SORD are presented in Figures 4 and 5. Multimedia Appendices 3 and 4 illustrate Bland-Altman for the other models. Percentage values are presented in the text (see Figures 4 and 5 and Multimedia Appendices 3 and 4 for true values).

Figure 4. Bland-Altman plot comparing seconds of sedentary behavior, standing, and walking between direct observation and activPAL against the Sedentary behaviOR Detector (SORD) activity tracker (model 1).
Mean differences (biases) between SORD model 1 and direct observation were 6.4% for sedentary (LoA=−4.3% to 17.1%), −8.7% for standing (LoA=−23.5% to 6.1%), and −8.9% for walking (LoA=−22.2% to 4.4%). Results of model 1 show wide limits, although the mean biases were below 10% for all activities. Relative to total activity durations, mean biases between SORD model 1 and activPAL were −2.5% for sedentary (LoA=−15.0% to 9.9%), 1.7% for standing (LoA=−23.3% to 26.9%), and 7.4% for walking (LoA=−2.3 to 17.1%). Results of model 1 comparing SORD to activPAL show wide limits.

Mean biases between SORD model 2 and direct observation were 0.3% for sitting and reclining (LoA=−0.3% to 0.9%), 1.19% for standing (LoA=−1.05% to 3.42%), and −4.71% for walking (LoA=−9.26% to −0.16%). Model 2 showed the narrowest LoA for “sitting and reclining,” standing, and walking, denoting excellent agreement with direct observation. All the mean biases were within ±10%. Relative to total activity durations, mean biases between SORD Model 2 and activPAL were −3.45% for sitting and reclining (LoA=−11.59% to 4.68%), 7.45% for standing (LoA=−5.04% to 19.95%), and −5.40% for walking (LoA=−11.44% to 0.64%). Results of model 2 comparing SORD to activPAL show a wider LoA, although mean biases are relatively low for “sitting and reclining” and walking.

Mean biases between SORD model 3 and direct observation were −6.4% for sitting (LoA=−18.6% to 5.7%), 12.4% for standing (LoA=−6.6% to 31.5%), and −4.9% for walking (LoA=−12.5% to 2.5%). Therefore, sitting and walking were
overestimated, while standing was underestimated. The mean bias was acceptable for sitting and walking but not standing. A narrow LoA were observed for walking.

Mean biases between SORD model 4 and direct observation were 52.2% for lying (LoA=–6.9% to 111.4%), –32.0% for sitting and reclining (LoA=–78.2% to 14.1%), 48.8% for standing (LoA=–13.7% to 111.4%), and –2.2% for walking (LoA=–6.7% to 2.2%). Therefore, “sitting and reclining” and walking were underestimated while lying and standing were overestimated. Model 4 shows the broadest LoA for “sitting and reclining” and standing, while the narrowest LoA were observed for walking in this model.

Discussion

This laboratory-based study assessed the criterion and convergent validity of a prototype activity tracker (ie, SORD). A high level of accuracy in detecting sitting, standing, and walking for the SORD device among adults was confirmed. Based on the Bland-Altman plots, high levels of agreement with direct observation demonstrated high criterion validity.

ActivPAL is a triaxial accelerometer that has been validated for detecting sitting, standing, and walking activity [29,41,42] and has been widely used in previous intervention studies [43-47]. However, a recent review found that activPAL has lower accuracy during fidgeting [48]. In this study, the agreement between SORD and activPAL was not ideal. The discrepancy observed might result from the inclusion of various fidgeting states. In addition, since activPAL does not enable real-time transmission of data to external devices or networks [29], it cannot be used for real-time or adaptive interventions. SitFit [30] is among the few devices that provide real-time feedback on SB. SitFit (PAL Technologies Ltd) is a pocket-worn device that requires appropriate clothing (eg, trousers with a front pocket), which is a barrier to its usability [30]. SitFit has an embedded screen to provide visual feedback to users and is also Bluetooth-enabled for connectivity to smartphones, tablets, and PCs. However, outputs generated by SitFit include sedentary time (sitting or lying), upright time, and step count [30]. The upright time includes both quiet standing and stepping [30], meaning that SitFit alone is not suitable for measuring standing as an outcome. Measuring standing and its variations (eg, fidgeting while standing) in real time will enable future intervention studies to identify distinct behavioral determinants of standing and to study its long-term clinical implications. As described in this study, SORD accurately measures sedentary (sitting and reclining), standing, and walking time. Other deep learning models (eg, model 4) examined whether the algorithms could distinguish lying from other sedentary states. A lower accuracy was observed for SORD in distinguishing lying from other sedentary activity states. Since the thigh is horizontal during lying posture, distinguishing sitting and lying postures with thigh-worn devices would be difficult. Methods that include rotational angle thresholds to determine the orientation of the thigh have been able to distinguish lying from sitting [49], even though these techniques require validation against direct observation to produce robust evidence.

A strength of this study is the inclusion of several variations of activity states (eg, sitting with outstretched legs, sitting while ankle-on-knee, and standing while shoulder on the wall), allowing more robust testing of the device accuracy and improving the generalizability of findings. For example, detecting standing as it appears in real-life situations and distinguishing from walking will enable the design of interventions measuring standing as a behavioral or clinical outcome. There are also limitations with this study, including the laboratory-based nature of the study. As with any laboratory-based experiment, it is possible that participants behave differently (eg, sit tall and neat and not as they would do normally). Moreover, a comparison between devices in terms of walking intensities was not conducted. This work is the first step in the validation of SORD, and longer-term studies in free-living environments would be necessary future steps to assess its practicality and accuracy under diverse conditions. The majority of participants in this study were female, and that might be considered a source of bias, that is, sex bias. However, evidence suggests that there are no significant differences between female individuals and male individuals in terms of posture, including sitting, standing, and walking [50]. Most participants were younger adults, and therefore the findings may not be generalizable to older adults. Investigating the usability of SORD in populations beyond young adults can help determine its broader applicability. Finally, we observed errors in the raw data from 2 participants for SORD and 3 others for activPAL.

In this study, we did not intend to compare or advance the activity recognition models; rather, the goal was to use the best approach for real-world applications of SORD for real-time intervention. The future development of SORD will include exploring other models (eg, ensemble learning).

In conclusion, SORD accurately detected sitting, standing, and walking activities among healthy young adults, and measurement accuracy was excellent compared to direct observation. While the current iteration of SORD displays promising levels of accuracy, it requires more work and real-world testing in an intervention to assess its applicability. Therefore, SORD holds potential for future integration into evidence- and theory-driven, real-time adaptive interventions to promote activity and reduce sedentary time.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental Figure 1. Confusion matrix for Model 3 classification algorithms. Sitting, standing and walking were included in the model.

[PNG File, 231 KB - formative_v811e47157_app1.png]

Multimedia Appendix 2
Supplemental Figure 2. Confusion matrix for Model 4 classification algorithms. Lying, sitting, reclining, standing and walking were included in the model.

[PNG File, 352 KB - formative_v811e47157_app2.png]

Multimedia Appendix 3
Supplemental Figure 3. Bland-Altman plot comparing seconds of sitting, standing and walking between the direct observation and SORD activity tracker (Model 3).

[PNG File, 99 KB - formative_v811e47157_app3.png]

Multimedia Appendix 4
Supplemental Figure 4. Bland-Altman plot comparing seconds of lying, sitting, reclining, standing and walking between the direct observation and SORD activity tracker (Model 4).

[PNG File, 129 KB - formative_v811e47157_app4.png]

References


34. Physical Activity Readiness Questionnaire PAR-Q and YOU. CSEP. 2002. URL: https://www.ons.org/sites/default/files/par-q.pdf [accessed 2023-12-12]


Abbreviations

- **LoA**: limits of agreement
- **PA**: physical activity
- **SB**: sedentary behavior
- **SORD**: Sedentary behaviOR Detector
Background: During the COVID-19 outbreak, video appointments became a popular method for health care delivery, particularly in the early stages of the pandemic. Although Mayo Clinic aimed to reduce face-to-face (F2F) appointments to prevent the spread of the virus, some patients continued seeing their health care providers in person. In the later stages of the pandemic, many patients became comfortable with video appointments, even if they were initially hesitant. However, a subset of patients continued to avoid video appointments. It is not yet clear what sociodemographic factors may be associated with this group of patients.

Objective: This cross-sectional study aimed to examine demographic and social determinant of health (SDoH) factors associated with persistent nonusers of video appointments among a sample of patients within a multistate health care organization. We also explored patient beliefs about the use of video for health care appointments.

Methods: We conducted a 1-time cross-sectional paper survey, mailed between July and December 2022, of patients matching the eligibility criteria: (1) aged ≥18 years as of April 2020, (2) Mayo Clinic Midwest, Florida, or Arizona patient, (3) did not use video appointment services during April-December 2020 but attended F2F appointments in the departments of primary care and psychiatry/psychology. The survey asked patients, “Have you ever had a video appointment with a healthcare provider?” “Yes” respondents were defined as “users” (adapted to video appointments), and “no” respondents were defined as “persistent nonusers” of video appointments. We analyzed demographics, SDoH, and patient beliefs toward video appointments in 2 groups: persistent nonusers of video appointments and users. We used chi-square and 2-tailed t tests for analysis.

Results: Our findings indicate that patients who were older, lived in rural areas, sought care at Mayo Clinic Midwest, and did not have access to the patient portal system were likely to be persistent nonusers of video appointments. Only 1 SDoH factor (not having a disability, handicap, or chronic disease) was associated with persistent nonuse of video appointments. Persistent nonusers of video appointments held personal beliefs such as discomfort with video communication, difficulty interpreting nonverbal cues, and personal preference for F2F appointments over video.

Conclusions: Our study identified demographic (older age and rural residence), sociodemographic factors (not having a disability, handicap, or chronic disease), and personal beliefs associated with patients’ decisions to choose between video versus F2F appointments for health care delivery. Health care institutions should assess patients’ negative attitudes toward technology prior to introducing them to digital health care services. Failing to do so may result in its restricted usage, negative patient experience,
and wasted resources. For patients who hold negative beliefs about technology but are willing to learn, a “digital health coordinator” could be assigned to assist with various digital health solutions.

**Introduction**

Since the COVID-19 pandemic, video appointments have been widely implemented for remotely delivered health care [1,2]. Both prepandemic and pandemic literature suggest that video visits improve provider access to patients, reduce patient travel and wait times, and provide health care quality comparable to face-to-face (F2F) appointments [3-6]. Despite these positive associations with telemedicine, video appointments for nonemergent care have not been as widely used by patients as expected. Studies show that individuals who could not adapt to digital health care delivery have faced significant health care access barriers during and since the pandemic [7]. In the current post-COVID-19 era, digital health care services are a new standard of care [8], and patients who need to be connected to the health care system digitally but are struggling to adapt to telemedicine may experience suboptimal health care [9]. Therefore, factors associated with nonengagement with video visits, especially in patients who have persistently not engaged in video appointments, require further exploration. While patients’ attitudes to telemedicine, especially in the COVID-19 era [10-12], have been explored, gaps remain in understanding social and individual characteristics associated with the persistent nonuse of video appointments for health care.

A large body of evidence suggests that older age, low education, poor digital access (broadband [BB] internet and smart devices), [13] and personal preferences [14] are independently and interactively associated with lower engagement with digital health care [14-17]. This is ironic, given that a critical reason behind the embarkation of digital health care technology was to provide uninterrupted health care access to those who live in remote areas where access to health care providers is limited, those who experience low socioeconomic status and associated transportation challenges and those with poor mobility due to old age and other constraints [18,19]. Evidence also shows that if the individual digital barriers are addressed [20-22], people are willing to engage in technology and participate in telemedicine programs. Preliminary public and institutional efforts to mitigate patient-related barriers to telemedicine are in their infancy but may include brief verbal and printed technology instructions, digital navigation programs for those who have poor digital literacy (comfort and ease of using technology), use of public Wi-Fi and “to-go kits” (smart devices with written instructions on connecting for a visit) [18,23]. Despite nationwide initiatives that accelerated after COVID-19 to encourage patients to use digital health care, many still chose to engage in F2F appointments [14,24]. This phenomenon was observed with and without social distancing associated with the COVID-19 pandemic.

Social determinants of health (SDoH) such as economic stability, access to quality education and health care, neighborhood safety and housing, community and social contexts, and experiences of racism and discrimination [25] significantly impact people’s well-being. The adverse outcomes associated with SDoH inequitably impact marginalized groups and prevent them from accessing quality health care. SDoH not only impact how easily and efficiently people can access health care, but also how they will access it (video vs F2F vs both).

For example, evidence shows that patients who identified as Black, indigenous, or people of color, and were non-English speaking patients and lived in neighborhoods with low socioeconomic status were less likely to engage in digital health care [17,26]. However, it has not been investigated yet which SDoH factor is linked to the persistent use and nonuse of video technology for health care appointments.

This cross-sectional study aimed to examine demographic and SDoH factors associated with no video use (self-reported persistent nonusers of video appointments) among a sample of patients within a large multistate health care organization. We also explored patient beliefs about the use of video for health care appointments. We hypothesized that certain demographic factors, including older age, being a woman, low education, rural residence, and SDoH, such as financial constraints and limited transportation options, may be associated with the persistent nonuse of video appointments.

**Methods**

**Setting**

Mayo Clinic consists of a large academic medical center and associated health system spanning the United States in 5 states (Minnesota, Wisconsin, Iowa, Florida, and Arizona). Mayo Clinic’s main campuses are located in Rochester, Minnesota; Phoenix, Arizona; and Jacksonville, Florida. Mayo Clinic Health System (MCHS) consists of clinics, hospitals, and other health care facilities in 4 regions in southern Minnesota, western Wisconsin, and northern Iowa. The Mayo Clinic Midwest (MN, WI, and IA) serves patients that are predominantly White, older people, and living in rural areas. In contrast, Mayo Clinic in Florida and Arizona serve a more diverse patient population.

**Ethics Approval**

This study was approved by the Mayo Clinic Institutional Review Board (21-004523).

**Study Overview and Design**

During earlier phases of COVID-19, in early 2020, the US government and the Centers for Disease Control and Prevention...
recommended social distancing measures, including stay-at-home orders and video appointments with health care providers [27]. Despite Mayo Clinic’s attempts to minimize F2F appointments to prevent the spread of the virus, many patients requested F2F appointments with their health providers. We were therefore interested in examining whether demographic and SDoH, including area-based metrics (where patients live), were associated with F2F visits. This study used a cross-sectional design with data collected from a 1-time survey administered to Mayo Clinic and MCHS patients.

**Survey Instrument and Measures**

The survey was designed using results from a prior qualitative study detailed elsewhere [14]. Guided by the qualitative results and informed by a scoping literature review, the survey items were developed to address existing gaps in the literature. The finalized paper survey was pretested with study staff with an estimated 10-15 minutes to complete. The survey included 21 items querying patient’s digital access such as BB internet connection and smart devices, digital literacy (the ease and comfort of using digital technology), use of the patient portal (Mayo Clinic patient online messaging system), use of video appointments, attitudes, and beliefs toward F2F versus video appointments and barriers to engaging in video appointments. The SDoH-related questions included in our survey were adapted from the Social Needs Screening Tool [28] (Multimedia Appendix 1).

Demographic characteristics (age, gender, and race or ethnicity), education status (highest during this study’s period), patient portal status (yes or no), and residence zip codes were extracted from the electronic health record (EHR). Rurality was ascertained from patient zip codes to identify corresponding rural-urban commuting area (RUCA) codes based on the University of Washington classification C method classification [29].

The dependent (outcome) variable was a dichotomous response (yes or no) to the question, “Have you ever had a video appointment with a healthcare provider?”

**Data Collection or Procedure**

We extracted data from the EHR of adult patients with this study’s eligibility of (1) being aged ≥18 years as of April 2020, (2) being a Mayo Clinic Midwest (Rochester or MCHS), Florida or Arizona patient, (3) not using video appointment services during the time frame of April-December 2020 but attending F2F appointments for nonemergent outpatient clinical care in the departments of primary care and psychiatry/psychology.

The Mayo Clinic Survey Research Center mailed eligible patients a survey in a prelabeled return envelope in early July 2022. By that time, a significant number of patients were oriented and made aware of video appointment procedures through self-learning and efforts by our health care institutions. Thus, the following survey item: “Have you ever had a video appointment with a healthcare provider?” with dichotomous responses “Yes/No,” provided valuable cross-sectional information distinguishing patients in this cohort in terms of their ability to adapt or not to evolving remote health care delivery appointments through video appointments for nonemergent care after April 2020. The respondents who marked “no” were defined as “persistent nonusers.” In contrast, those who responded “yes” were defined as “users” who, despite not having used video appointments between April and December 2020, adapted to the changing digital landscape, using them later.

Surveys were mailed to eligible patients stratified by departmental visit type (psychiatry/psychology versus primary care), demographic characteristics (gender, race, and Mayo Clinic location), and if the patient has an active patient portal account. The Survey Research Center mailed reminder letters to nonresponders in August 2022 and then conducted phone call reminders to nonresponders in October-December 2022. Survey participation was closed in January 2023. Survey respondents received a sheet of forever stamps valued at US $5.

**Statistical Analysis**

Demographics, SDoH, and patient beliefs about video encounters were compared across groups, persistent nonusers of video appointments, and users groups, using the chi-square (exact) test for categorical variables and the 2-sample t test (rank sum) for continuous variables. In all cases, $P$ values <.05 were considered statistically significant.

**Results**

**Overview**

Respondent sociodemographic characteristics (N=321) are described in Table 1 overall and by use of video appointments. The survey response rate was 11% (321/3000). In the total respondent sample, 172 (54%) were women, 217 (68%) were White, 169 (53%) had bachelor’s or higher education degrees (persistent nonusers vs users; 84, 52.5% vs 85, 52.8%), and 282 (88%) were urban dwelling (persistent nonusers vs users; 133, 83.1% vs 149, 92.5%; $P=0.01$). In addition, 266 (83%) had access to an online patient portal account (persistent nonusers vs users; 122, 76.2% vs 144, 89.4%; $P=0.002$).
Table 1. Demographic factors associated with the using and not using video appointments.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Total (N=321), n (%)</th>
<th>Persistent nonusers of video (n=160), n (%)</th>
<th>Users of video (n=161), n (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (y)</strong></td>
<td></td>
<td></td>
<td></td>
<td>.001&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>57.440 (16.400)</td>
<td>60.434 (16.849)</td>
<td>54.465 (15.425)</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>18.111-90.010</td>
<td>18.111-90.010</td>
<td>21.791-86.867</td>
<td></td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td>.05&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Women</td>
<td>172 (53.6)</td>
<td>77 (48.1)</td>
<td>95 (59)</td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>149 (46.4)</td>
<td>83 (51.9)</td>
<td>66 (41)</td>
<td></td>
</tr>
<tr>
<td><strong>Specialty</strong></td>
<td></td>
<td></td>
<td></td>
<td>&gt;.99&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Community medicine</td>
<td>309 (96.3)</td>
<td>154 (96.2)</td>
<td>155 (96.3)</td>
<td></td>
</tr>
<tr>
<td>General internal medicine</td>
<td>6 (1.9)</td>
<td>3 (1.9)</td>
<td>3 (1.9)</td>
<td></td>
</tr>
<tr>
<td>Psychiatry and psychology</td>
<td>6 (1.9)</td>
<td>3 (1.9)</td>
<td>3 (1.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Site</strong></td>
<td></td>
<td></td>
<td></td>
<td>.02&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Arizona</td>
<td>134 (41.7)</td>
<td>67 (41.9)</td>
<td>67 (41.6)</td>
<td></td>
</tr>
<tr>
<td>Florida</td>
<td>96 (29.9)</td>
<td>38 (23.8)</td>
<td>58 (36)</td>
<td></td>
</tr>
<tr>
<td>Mayo Clinic Midwest</td>
<td>91 (28.3)</td>
<td>55 (34.4)</td>
<td>36 (22.4)</td>
<td></td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
<td>.38&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Non-White</td>
<td>89 (27.7)</td>
<td>47 (29.4)</td>
<td>42 (26.1)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>15 (4.7)</td>
<td>5 (3.1)</td>
<td>10 (6.2)</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>217 (67.6)</td>
<td>108 (67.5)</td>
<td>109 (67.7)</td>
<td></td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td>.54&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Less than equal to 12th grade</td>
<td>30 (9.3)</td>
<td>16 (10)</td>
<td>14 (8.7)</td>
<td></td>
</tr>
<tr>
<td>Some college, no degree</td>
<td>22 (6.9)</td>
<td>9 (5.6)</td>
<td>13 (8.1)</td>
<td></td>
</tr>
<tr>
<td>Associate degree</td>
<td>36 (11.2)</td>
<td>18 (11.2)</td>
<td>18 (11.2)</td>
<td></td>
</tr>
<tr>
<td>Bachelors</td>
<td>84 (26.2)</td>
<td>36 (22.5)</td>
<td>48 (29.8)</td>
<td></td>
</tr>
<tr>
<td>Higher education</td>
<td>85 (26.5)</td>
<td>48 (30)</td>
<td>37 (23)</td>
<td></td>
</tr>
<tr>
<td>Decline to answer</td>
<td>64 (19.9)</td>
<td>33 (20.6)</td>
<td>31 (19.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td></td>
<td></td>
<td>.36&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Married</td>
<td>238 (74.1)</td>
<td>117 (73.1)</td>
<td>121 (75.2)</td>
<td></td>
</tr>
<tr>
<td>Single, separate, divorced, or widowed</td>
<td>71 (22.1)</td>
<td>39 (24.4)</td>
<td>32 (19.9)</td>
<td></td>
</tr>
<tr>
<td>Unknown or chose “N”</td>
<td>12 (3.7)</td>
<td>4 (2.5)</td>
<td>8 (5)</td>
<td></td>
</tr>
<tr>
<td><strong>Rural versus urban</strong></td>
<td></td>
<td></td>
<td></td>
<td>.01&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Rural</td>
<td>39 (12.1)</td>
<td>27 (16.9)</td>
<td>12 (7.5)</td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>282 (87.9)</td>
<td>133 (83.1)</td>
<td>149 (92.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Portal (online patient messaging system)</strong></td>
<td></td>
<td></td>
<td></td>
<td>.002&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>No</td>
<td>55 (17.1)</td>
<td>38 (23.8)</td>
<td>17 (10.6)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>266 (82.9)</td>
<td>122 (76.2)</td>
<td>144 (89.4)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Two-sample 2-tailed t test.
<sup>b</sup>Chi-squared test.
<sup>c</sup>Fisher exact test.
Demographic Correlates to Persistent Nonuse of Video Appointments

Persistent nonusers of video appointments were older than users ($P=.001$). In addition, patients living in rural residences ($P=.01$) were more likely to be persistent nonusers of video appointments. Other demographic factors, such as gender, education, and race, were not significantly different between persistent nonusers and users of video appointments (Table 1).

Institution Site Correlates to Persistent Nonuse of Video Appointments

Patients who sought care at Mayo Clinic Midwest, comprising Mayo Clinic, Rochester, and MCHS, were more likely to be persistent nonusers of video appointments ($P=.02$; Table 1).

Social Correlates to Persistent Nonuse of Video Appointments

Patients without any disability, handicap, or chronic disease were more likely to be persistent nonusers of video appointments than users ($P=.01$; Table 2). Other SDoH-related factors were not statistically significant.
Table 2. Social determinant of health factors associated with using and not using video appointments.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (N=321)</th>
<th>Persistent nonusers of video (n=160)</th>
<th>Users of video (n=161)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within the past 12 months, did you worry that your food would run out before you got money to buy more?</strong></td>
<td></td>
<td></td>
<td></td>
<td>.12</td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>No, n (%)</td>
<td>311 (97.2)</td>
<td>157 (98.7)</td>
<td>154 (95.7)</td>
<td></td>
</tr>
<tr>
<td>Yes, n (%)</td>
<td>7 (2.2)</td>
<td>1 (0.6)</td>
<td>6 (3.7)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer, n (%)</td>
<td>2 (0.6)</td>
<td>1 (0.6)</td>
<td>1 (0.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Within the past 12 months, did the food you bought just not last, and you did not have money to get more?</strong></td>
<td></td>
<td></td>
<td></td>
<td>.45</td>
</tr>
<tr>
<td>Missing</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>No, n (%)</td>
<td>311 (97.5)</td>
<td>156 (98.7)</td>
<td>155 (96.3)</td>
<td></td>
</tr>
<tr>
<td>Yes, n (%)</td>
<td>7 (2.2)</td>
<td>2 (1.3)</td>
<td>5 (3.1)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer, n (%)</td>
<td>1 (0.3)</td>
<td>0 (0)</td>
<td>1 (0.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Do you have housing?</strong></td>
<td></td>
<td></td>
<td></td>
<td>.10</td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>No, n (%)</td>
<td>9 (2.8)</td>
<td>7 (4.4)</td>
<td>2 (1.2)</td>
<td></td>
</tr>
<tr>
<td>Yes, n (%)</td>
<td>310 (96.9)</td>
<td>152 (95.6)</td>
<td>158 (98.1)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer, n (%)</td>
<td>1 (0.3)</td>
<td>0 (0)</td>
<td>1 (0.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Are you worried about losing your housing?</strong></td>
<td></td>
<td></td>
<td></td>
<td>&gt;.99</td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>No, n (%)</td>
<td>309 (96.6)</td>
<td>154 (96.9)</td>
<td>155 (96.3)</td>
<td></td>
</tr>
<tr>
<td>Yes, n (%)</td>
<td>9 (2.8)</td>
<td>4 (2.5)</td>
<td>5 (3.1)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer, n (%)</td>
<td>2 (0.6)</td>
<td>1 (0.6)</td>
<td>1 (0.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Within the past 12 months, have you or your family members you live with been without utilities?</strong></td>
<td></td>
<td></td>
<td></td>
<td>.62</td>
</tr>
<tr>
<td>Missing</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>No, n (%)</td>
<td>311 (98.4)</td>
<td>155 (99.4)</td>
<td>156 (97.5)</td>
<td></td>
</tr>
<tr>
<td>Yes, n (%)</td>
<td>4 (1.3)</td>
<td>1 (0.6)</td>
<td>3 (1.9)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer, n (%)</td>
<td>1 (0.3)</td>
<td>0 (0)</td>
<td>1 (0.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Within the past 12 months, lack of transportation?</strong></td>
<td></td>
<td></td>
<td></td>
<td>.28</td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>No, n (%)</td>
<td>312 (97.5)</td>
<td>154 (96.9)</td>
<td>158 (98.1)</td>
<td></td>
</tr>
<tr>
<td>Yes, n (%)</td>
<td>7 (2.2)</td>
<td>5 (3.1)</td>
<td>2 (1.2)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer, n (%)</td>
<td>1 (0.3)</td>
<td>0 (0)</td>
<td>1 (0.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Within the past 12 months, did you have trouble paying your bills?</strong></td>
<td></td>
<td></td>
<td></td>
<td>.89</td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>No, n (%)</td>
<td>304 (95)</td>
<td>152 (95.6)</td>
<td>152 (94.4)</td>
<td></td>
</tr>
<tr>
<td>Yes, n (%)</td>
<td>14 (4.4)</td>
<td>6 (3.8)</td>
<td>8 (5)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer, n (%)</td>
<td>2 (0.6)</td>
<td>1 (0.6)</td>
<td>1 (0.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Does any disability, handicap, or chronic disease make it difficult for you to engage in your typical activities?</strong></td>
<td></td>
<td></td>
<td></td>
<td>.01</td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>No, n (%)</td>
<td>279 (87.2)</td>
<td>147 (92.5)</td>
<td>132 (82)</td>
<td></td>
</tr>
<tr>
<td>Yes, n (%)</td>
<td>36 (11.2)</td>
<td>11 (6.9)</td>
<td>25 (15.5)</td>
<td></td>
</tr>
<tr>
<td>Prefer not to answer, n (%)</td>
<td>5 (1.6)</td>
<td>1 (0.6)</td>
<td>4 (2.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Are you currently working for pay?</strong></td>
<td></td>
<td></td>
<td></td>
<td>.61</td>
</tr>
</tbody>
</table>
Video Encounter-Related Beliefs Correlate to Persistent Nonusers of Video Appointments

Scenario 1: “Imagine you are having a video appointment with a Mayo Clinic doctor for a general medicine health check-up that does not require any procedures or exams. Further, imagine you have seen this doctor before for a face-to-face or in-person visit.” A significantly lower proportion of persistent nonusers responded “agree,” while a significantly higher proportion of persistent nonusers of video appointments responded “somewhat disagree” or “disagree,” respectively, to the following statements in response to this scenario: “I am confident my doctor would be able to address any medical concerns effectively” ($P=0.047$), “I am confident I would be able to express all my concerns clearly” ($P=0.04$) and “I am confident I would feel comfortable enough to talk openly” ($P<0.001$) compared to users (Table 3). No other responses were statistically significantly associated with the comparison groups.
Table 3. Patients’ beliefs about video encounters and their association with the use of video appointments.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (N=321)</th>
<th>Persistent nonusers of video (n=160)</th>
<th>Users of video (n=161)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am confident my doctor would be able to address any medical concerns effectively</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=miss</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>.047a</td>
</tr>
<tr>
<td>1=agree, n (%)</td>
<td>186 (58.7)</td>
<td>80 (51)</td>
<td>106 (66.2)</td>
<td></td>
</tr>
<tr>
<td>2=somewhat agree, n (%)</td>
<td>91 (28.7)</td>
<td>52 (33.1)</td>
<td>39 (24.4)</td>
<td></td>
</tr>
<tr>
<td>3=somewhat disagree, n (%)</td>
<td>19 (6)</td>
<td>12 (7.6)</td>
<td>7 (4.4)</td>
<td></td>
</tr>
<tr>
<td>4=disagree, n (%)</td>
<td>21 (6.6)</td>
<td>13 (8.3)</td>
<td>8 (5)</td>
<td></td>
</tr>
<tr>
<td>I am confident I would be able to express all my concerns clearly</td>
<td></td>
<td></td>
<td></td>
<td>.04b</td>
</tr>
<tr>
<td>N=miss</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1=agree, n (%)</td>
<td>222 (70.3)</td>
<td>99 (63.1)</td>
<td>123 (77.4)</td>
<td></td>
</tr>
<tr>
<td>2=somewhat agree, n (%)</td>
<td>64 (20.3)</td>
<td>41 (26.1)</td>
<td>23 (14.5)</td>
<td></td>
</tr>
<tr>
<td>3=somewhat disagree, n (%)</td>
<td>16 (5.1)</td>
<td>9 (5.7)</td>
<td>7 (4.4)</td>
<td></td>
</tr>
<tr>
<td>4=disagree, n (%)</td>
<td>14 (4.4)</td>
<td>8 (5.1)</td>
<td>6 (3.8)</td>
<td></td>
</tr>
<tr>
<td>I am confident I would feel comfortable enough to talk openly</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001b</td>
</tr>
<tr>
<td>N=miss</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1=agree, n (%)</td>
<td>247 (78.4)</td>
<td>106 (67.9)</td>
<td>141 (88.7)</td>
<td></td>
</tr>
<tr>
<td>2=somewhat agree, n (%)</td>
<td>46 (14.6)</td>
<td>33 (21.2)</td>
<td>13 (8.2)</td>
<td></td>
</tr>
<tr>
<td>3=somewhat disagree, n (%)</td>
<td>11 (3.5)</td>
<td>8 (5.1)</td>
<td>3 (1.9)</td>
<td></td>
</tr>
<tr>
<td>4=disagree, n (%)</td>
<td>11 (3.5)</td>
<td>9 (5.8)</td>
<td>2 (1.3)</td>
<td></td>
</tr>
<tr>
<td>I feel video appointments should cost the same and are of equal value to face-to-face appointments</td>
<td></td>
<td></td>
<td></td>
<td>.09a</td>
</tr>
<tr>
<td>N=miss</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>1=agree, n (%)</td>
<td>66 (21)</td>
<td>25 (16)</td>
<td>41 (25.9)</td>
<td></td>
</tr>
<tr>
<td>2=somewhat agree, n (%)</td>
<td>80 (25.5)</td>
<td>45 (28.8)</td>
<td>35 (22.2)</td>
<td></td>
</tr>
<tr>
<td>3=somewhat disagree, n (%)</td>
<td>91 (29)</td>
<td>43 (27.6)</td>
<td>48 (30.4)</td>
<td></td>
</tr>
<tr>
<td>4=disagree, n (%)</td>
<td>77 (24.5)</td>
<td>43 (27.6)</td>
<td>34 (21.5)</td>
<td></td>
</tr>
</tbody>
</table>

Scenario #1: “Imagine you are having a video appointment with a Mayo Clinic doctor for a general medicine health check-up that does not require any procedures or exams. Further, imagine you have seen this doctor before for a face-to-face or in-person visit.”

Scenario #2: Imagine you are having an appointment with a Mayo Clinic psychiatrist or psychologist that does not require any procedures or exams. Further, imagine you have seen this doctor before for a face-to-face or in-person visit.

I am confident my doctor would be able to address any medical concerns effectively

I am confident I would be able to express all my concerns clearly

I am confident I would feel comfortable enough to talk openly

https://formative.jmir.org/2024/1/e50572
<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (N=321)</th>
<th>Persistent nonusers of video (n=160)</th>
<th>Users of video (n=161)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N=miss</td>
<td>9</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1=agree, n (%)</td>
<td>202 (64.7)</td>
<td>84 (54.2)</td>
<td>118 (75.2)</td>
<td>.26a</td>
</tr>
<tr>
<td>2= somewhat agree, n (%)</td>
<td>67 (21.5)</td>
<td>42 (27.1)</td>
<td>25 (15.9)</td>
<td></td>
</tr>
<tr>
<td>3= somewhat disagree, n (%)</td>
<td>24 (7.7)</td>
<td>15 (9.7)</td>
<td>9 (5.7)</td>
<td></td>
</tr>
<tr>
<td>4= disagree, n (%)</td>
<td>19 (6.1)</td>
<td>14 (9)</td>
<td>5 (3.2)</td>
<td></td>
</tr>
<tr>
<td>I feel video appointments should cost the same and are of equal value to face-to-face appointments</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=miss</td>
<td>9</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1=agree, n (%)</td>
<td>94 (30.1)</td>
<td>39 (25.2)</td>
<td>55 (35)</td>
<td>.26a</td>
</tr>
<tr>
<td>2= somewhat agree, n (%)</td>
<td>84 (26.9)</td>
<td>46 (29.7)</td>
<td>38 (24.2)</td>
<td></td>
</tr>
<tr>
<td>3= somewhat disagree, n (%)</td>
<td>70 (22.4)</td>
<td>38 (24.5)</td>
<td>32 (20.4)</td>
<td></td>
</tr>
<tr>
<td>4= disagree, n (%)</td>
<td>64 (20.5)</td>
<td>32 (20.6)</td>
<td>32 (20.4)</td>
<td></td>
</tr>
<tr>
<td>Video encounter-related beliefs not specific to any discipline</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am confident I would be able to understand when the doctor explains my symptoms or health</td>
<td></td>
<td></td>
<td></td>
<td>.046b</td>
</tr>
<tr>
<td>N=miss</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1=agree, n (%)</td>
<td>206 (65.6)</td>
<td>90 (58.1)</td>
<td>116 (73)</td>
<td></td>
</tr>
<tr>
<td>2= somewhat agree, n (%)</td>
<td>79 (25.2)</td>
<td>47 (30.3)</td>
<td>32 (20.1)</td>
<td></td>
</tr>
<tr>
<td>3= somewhat disagree, n (%)</td>
<td>17 (5.4)</td>
<td>10 (6.5)</td>
<td>7 (4.4)</td>
<td></td>
</tr>
<tr>
<td>4= disagree, n (%)</td>
<td>12 (3.8)</td>
<td>8 (5.2)</td>
<td>4 (2.5)</td>
<td></td>
</tr>
<tr>
<td>I am confident I would be able to read my doctor's facial expressions or nonverbal cues</td>
<td></td>
<td></td>
<td></td>
<td>.05a</td>
</tr>
<tr>
<td>N=miss</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1=agree, n (%)</td>
<td>140 (44.4)</td>
<td>57 (36.8)</td>
<td>83 (51.9)</td>
<td></td>
</tr>
<tr>
<td>2= somewhat agree, n (%)</td>
<td>114 (36.2)</td>
<td>62 (40)</td>
<td>52 (32.5)</td>
<td></td>
</tr>
<tr>
<td>3= somewhat disagree, n (%)</td>
<td>41 (13)</td>
<td>25 (16.1)</td>
<td>16 (10)</td>
<td></td>
</tr>
<tr>
<td>4= disagree, n (%)</td>
<td>20 (6.3)</td>
<td>11 (7.1)</td>
<td>9 (5.6)</td>
<td></td>
</tr>
<tr>
<td>I am confident I would be able to hear my doctor clearly</td>
<td></td>
<td></td>
<td></td>
<td>.004b</td>
</tr>
<tr>
<td>N=miss</td>
<td>7</td>
<td>6</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1=agree, n (%)</td>
<td>200 (63.7)</td>
<td>85 (55.2)</td>
<td>115 (71.9)</td>
<td></td>
</tr>
<tr>
<td>2= somewhat agree, n (%)</td>
<td>79 (25.2)</td>
<td>43 (27.9)</td>
<td>36 (22.5)</td>
<td></td>
</tr>
<tr>
<td>3= somewhat disagree, n (%)</td>
<td>22 (7)</td>
<td>16 (10.4)</td>
<td>6 (3.8)</td>
<td></td>
</tr>
<tr>
<td>4= disagree, n (%)</td>
<td>13 (4.1)</td>
<td>10 (6.5)</td>
<td>3 (1.9)</td>
<td></td>
</tr>
<tr>
<td>I would enjoy connecting with my doctor as much as as if the appointment were face-to-face</td>
<td></td>
<td></td>
<td></td>
<td>.009b</td>
</tr>
<tr>
<td>N=miss</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1=agree, n (%)</td>
<td>108 (34.1)</td>
<td>40 (25.5)</td>
<td>68 (42.5)</td>
<td></td>
</tr>
<tr>
<td>2= somewhat agree, n (%)</td>
<td>88 (27.8)</td>
<td>46 (29.3)</td>
<td>42 (26.2)</td>
<td></td>
</tr>
<tr>
<td>3= somewhat disagree, n (%)</td>
<td>68 (21.5)</td>
<td>38 (24.2)</td>
<td>30 (18.8)</td>
<td></td>
</tr>
<tr>
<td>4= disagree, n (%)</td>
<td>53 (16.7)</td>
<td>33 (21)</td>
<td>20 (12.5)</td>
<td></td>
</tr>
<tr>
<td>I would feel comfortable talking with a doctor I have met before in-person</td>
<td></td>
<td></td>
<td></td>
<td>.16b</td>
</tr>
<tr>
<td>N=miss</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1=agree, n (%)</td>
<td>221 (70.4)</td>
<td>101 (65.2)</td>
<td>120 (75.5)</td>
<td></td>
</tr>
<tr>
<td>2= somewhat agree, n (%)</td>
<td>64 (20.4)</td>
<td>36 (23.2)</td>
<td>28 (17.6)</td>
<td></td>
</tr>
<tr>
<td>3= somewhat disagree, n (%)</td>
<td>15 (4.8)</td>
<td>8 (5.2)</td>
<td>7 (4.4)</td>
<td></td>
</tr>
</tbody>
</table>
Principal Findings

This cross-sectional study demonstrated demographic and SDoH factors associated with persistent nonusers of video appointments for health care in a multisite medical institution. We observed that about 50% (161 of 321) of respondents persistently have not engaged with video appointments for non-emergent primary and psychiatric care since the start of video appointments in our institution (April 2020). We further observed that patients of older age, those living in rural residences, those who sought care at Mayo Clinic Midwest and those who did not have access to the patient portal system were more likely to be persistent nonusers of video appointments. Only a single SDoH-related factor (not having a disability, handicap, or chronic disease) was associated with persistent nonuse of video appointments. We also observed that individuals held certain personal beliefs about video appointments that were associated with their decision to use versus not use video appointments for health care. The persistent nonusers of video appointments held beliefs that included being potentially uncomfortable communicating with their doctor through video, difficulty reading their doctor’s facial expressions or nonverbal cues, struggle to hear the doctor clearly, and overall better comfort with F2F appointments over video appointments.

Much evidence has demonstrated that older patients have limited engagement with telemedicine, including using video appointments for their health care needs [30-32]. Our study observed an analogous association with older age correlated with persistent nonuse of video appointments for health care. Given that few individuals in our sample experienced substantial limitations in SDoH (Table 2) and that most respondents lived in urban dwellings and had access to the online patient portal (which requires smart devices and internet BB connection), we speculate that factors other than just digital access barriers should be considered when approaching older patients for increasing digital engagement. One possible factor is limited interest in digital health care due to negative personal beliefs toward video appointments. Given that older adulthood is a period when many individuals experience a decline in physical and cognitive health and could lose interest in exploring newer concepts (technology in this case), it is essential for health care providers and health care systems to take a patient-centered approach to understand the reasoning behind an older adult patient’s preference for in-person versus video appointment and provide the appropriate support and develop barrier mitigating strategies tailored to age to engage these individuals with needed care. This study found that patients who lived in rural areas were more likely to be persistent nonvideo users. This finding has been established by many studies. A key reason for the rural-urban digital health disparity is unequal access to BB connections. Additionally, individuals living in rural areas tend to be older, have limited education, and lack the financial resources to invest in BB connections and smartphones. Overall, our research effectively collected information on demographic factors associated with persistent nonusers of video appointments for health care in our institution.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (N=321)</th>
<th>Persistent nonusers of video (n=160)</th>
<th>Users of video (n=161)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>I would feel comfortable talking with a doctor I have never met before in-person</td>
<td>14 (4.5)</td>
<td>10 (6.5)</td>
<td>4 (2.5)</td>
<td>.01a</td>
</tr>
<tr>
<td>N=miss</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1=agree, n (%)</td>
<td>84 (26.8)</td>
<td>34 (21.9)</td>
<td>50 (31.4)</td>
<td></td>
</tr>
<tr>
<td>2=somewhat agree, n (%)</td>
<td>102 (32.5)</td>
<td>44 (28.4)</td>
<td>58 (36.5)</td>
<td></td>
</tr>
<tr>
<td>3=somewhat disagree, n (%)</td>
<td>76 (24.2)</td>
<td>44 (28.4)</td>
<td>32 (20.1)</td>
<td></td>
</tr>
<tr>
<td>4=disagree, n (%)</td>
<td>52 (16.6)</td>
<td>33 (21.3)</td>
<td>19 (11.9)</td>
<td></td>
</tr>
</tbody>
</table>

aChi-squared test.

bFisher exact test.

Scenario 2: “Imagine you are having an appointment with a Mayo Clinic Psychiatrist or Psychologist that does not require any procedures or exams. Further, imagine you have seen this doctor before for a face-to-face or in-person visit.” A significantly lower proportion of persistent nonusers of video appointments responded “agree,” while a significantly higher proportion of persistent nonusers of video appointments responded “somewhat disagree” or “disagree,” respectively, to the following statements in response to this scenario: “I am confident I would be able to express all my concerns clearly” (P=.03), and “I am confident I would feel comfortable enough to talk openly” (P=.001) compared to users. No other responses were statistically significantly associated with the comparison groups.

Video Encounter-Related Beliefs as a Correlate to Persistent Nonuse of Video Appointments

A significantly lower proportion of persistent nonusers of video appointments responded “agree,” while a significantly higher proportion of persistent nonusers of video appointments responded “somewhat disagree” or “disagree” to the following statements: “I am confident I would be able to understand when the doctor explains my symptoms/health” (P=.046), “I am confident I would be able to read my doctor’s facial expressions or non-verbal cues” (P=.05), “I am confident I would be able to hear my doctor clearly” (P=.004), “I would enjoy connecting with my doctor as much as if the appointment were face-to-face” (P=.009), and “I would feel comfortable talking with a doctor I have never met before in-person” (P=.1) compared to users. No other responses were statistically significantly associated with the comparison groups.

Discussion

Principal Findings

This cross-sectional study demonstrated demographic and SDoH factors associated with persistent nonusers of video appointments for health care in a multisite medical institution. We observed that about 50% (161 of 321) of respondents persistently have not engaged with video appointments for non-emergent primary and psychiatric care since the start of video appointments in our institution (April 2020). We further observed that patients of older age, those living in rural residences, those who sought care at Mayo Clinic Midwest and other medical institutions, and those who did not have access to the patient portal system were more likely to be persistent nonusers of video appointments. Only a single SDoH-related factor (not having a disability, handicap, or chronic disease) was associated with persistent nonuse of video appointments. We also observed that individuals held certain personal beliefs about video appointments that were associated with their decision to use versus not use video appointments for health care. The persistent nonusers of video appointments held beliefs that included being potentially uncomfortable communicating with their doctor through video, difficulty reading their doctor’s facial expressions or nonverbal cues, struggle to hear the doctor clearly, and overall better comfort with F2F appointments over video appointments.
indicators associated with not using video appointments that parallels the geographic demographics of Mayo Clinic, Rochester and various MCHS locations in rural areas with mostly older White residents.

As part of this study, we also aimed to evaluate the social factors or SDoH-related concerns that contributed to the persistent nonusers of video appointments. In our sample, the only factor associated with not using video appointments was not having a disability, handicap, or chronic disease. It is possible that their mobility or health allowed for greater flexibility in choosing an F2F visit or that they simply had fewer visits overall and were, therefore, less likely to choose video visits as an alternative. On the other hand, video appointments could be specifically beneficial for patients with disabilities or chronic illnesses who may have challenges with physical energy or mobility, be at higher risk for contracting illness when in public or have more health care appointments to attend overall. Given that there was limited variation in SDoH within our sample, this may have limited our ability to identify potential correlations between SDoH factors and selecting video or F2F appointments. Large-scale studies with socially diverse patients are required to fully understand the extent to which SDoH factors play in patients’ decision-making in choosing health care delivery methods. This understanding will further enhance patient outreach efforts and strengthen high-impact population health and research initiatives.

Since the pandemic, a significant public effort has been made to increase patients’ digital access at state, federal, and institutional levels [33]. Still, some patients may be unenthusiastic about attending video appointments [14,34]. We found that persistent nonusers of video appointments feel that they may not be able to “express” their concerns and are not able to “feel comfortable enough to talk openly” when having video appointments with primary care and psychiatry practicing physicians. Enjoying F2F encounters better than video appointments and potentially being unable to hear doctors clearly during a video appointment were 2 other beliefs that persistent nonusers of video appointments cited in high proportion. These patients conveyed these beliefs despite evidence of never engaging in video appointments in our institution. It is possible that individuals who consistently do not use video appointments have formed their opinions based on information obtained from sources other than their personal experience. These sources may include internet forums or the opinions of their peers. Another potential explanation could be poor digital experience when they attempted to engage in video appointments due to limited digital access (low-speed internet), language barrier, and low digital and health literacy [35]. It has been widely understood that poor digital experience could trigger patients’ anxiety regarding existing and emerging technology used in health care and may lead to its avoidance. This problem could be solved by appointing a “digital health coordinator” at the institutional level whose sole responsibility should be assisting patients with digital health solutions. This could overcome the perceived reluctance of patients to use digital services for health consultations. In addition, health care institutions should take into account the strong negative attitudes of this group toward video appointments when introducing them to digital technology for health care delivery.

Overall, our study results may inspire researchers to initiate a conversation about video adoption that goes beyond digital access and literacy. Our research examined the impact of SDoH and confidence or belief in video appointments adoption. Previous studies have not investigated which SDoH is most closely associated with video use. Furthermore, individuals with digital access and digital literacy may still choose to refrain from using video appointments. Therefore, objective measures should consider patient beliefs. Health care institutions should assess and evaluate patient preferences when implementing digital health care, especially those with digital competencies. We have yet to identify any digital literacy (validated) scales that have assessed an individual’s digital belief as one of the variables (negative vs positive) to assess overall digital literacy. From the perspective of behavior change theories [36], it is widely accepted that targeting one’s beliefs is essential for behavior change (digital adoption in this case). Therefore, our study adds novelty to the literature by informing researchers about understanding digital beliefs as a confounder in digital literacy and adoption. We suggest that through the community-based participatory research (CBPR) approach, researchers should attempt to identify facilitators to expedite behavior change. In a subsequent study with a larger sample size, it would be worth exploring if patients with limited interest in video appointments have sufficient digital access and literacy.

Limitations
Our study has several limitations, including the low survey response rate, which may have led to selection bias, resulting in a study population that does not accurately represent the target population, and respondents may differ systematically from nonrespondents. We used the self-reported data and the possibility of recall bias. To help alleviate such a concern, we did verify eligibility and the existence of an F2F appointment via EHR. Additionally, though we tried to enroll participants from diverse backgrounds, the majority of patients in our sample were White, lived in urban areas, and did not experience major social challenges, limiting the generalizability of our findings. Our demographic variables were not extensive due to lack of availability or missing values in the data extracted from EHR. In addition, the survey did not include factors related to the health care system, such as whether patients requested a video appointment, if video appointments were encouraged and offered to patients, or if video appointments were offered but declined by the patient. The results of our study may also lack generalizability because the sample was derived from Mayo Clinic patients and there was no feasible way to assess if patients sought care outside Mayo Clinic and used video visits. However, we enrolled patients who have their primary care providers (PCPs) at Mayo Clinic (ie, paneled patients), reducing the likelihood of video visits being done outside of our health care system. According to FAIR Health [37], a national database of private and Medicare claims data, only 0.1% of all claims nationally in 2019 were related to telehealth. This percentage was even lower in rural areas. Based on these data, it is highly unlikely for patients with a PCP at Mayo Clinic to seek outside video-based care. It’s important to note that FAIR Health data
include not just video visits, but also other telehealth technologies such as mobile health, remote patient monitoring, and store and forward technologies. We aimed to gather diverse data by including Mayo Clinic, Arizona, and Mayo Clinic, Florida. Despite a larger number of responses from Florida, the participants who responded were not from a diverse population. The studies show that there are differences in participation rates based on race, including lower rates of completing consent forms and responding to surveys, with Blacks and Hispanics being the most underrepresented [38]. Future work should explore the patterns of video use in more diverse patient populations, especially those who may be more likely to face barriers to health care (eg, patients living in rural areas and patients experiencing challenges with transportation). Our study also had several strengths, including that our sample was drawn from a multistate institution spanning rural and urban settings, the use of a validated measure of SDoH and the inclusion of scenario-based questions to better understand patients’ beliefs about video encounters.

Conclusions
Our study identified sociodemographic factors and personal beliefs about video appointments that influenced patients’ decisions to choose between video versus F2F appointments for health care delivery. The patients who were older, lived in rural residences, sought care at Mayo Clinic Midwest, and who did not have access to the patient portal were more likely to be persistent nonusers of video appointments. We observed a single SDoH factor, that is not having a disability, handicap, or chronic diseases associated with persistent nonusers of video appointments. Not being able to adequately “express” their medical concerns and not “feel comfortable enough to talk openly” and adequately listen to their provider were notable beliefs held by persistent nonusers of video appointments. We recommend that health care institutions consider and evaluate patients’ strong negative beliefs toward video appointments when introducing them to digital technology for health care delivery. Conducting large-scale studies that encompass a diverse range of social and demographic backgrounds is imperative to comprehend why patients prefer video or in-person appointments. Only through such research can we identify the factors that influence their decision-making process.

Acknowledgments
This work was supported by a Mayo Clinic Clinical Practice Committee Eradicating Racism Award and Clinical and Translational Science Awards (CTSA) from the National Center for Advancing Translational Science (grant UL1 TR002377). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health. The funding source had no role in this study’s design or in the collection, analysis, and interpretation of data; in writing the paper; or in the decision to submit the paper for publication. We thank the Mayo Clinic Survey Research Center for providing assistance with the survey design and data collection.

Authors’ Contributions
PS and CP conceptualized this study, performed the investigation, acquired funding, provided resources, designed the methodology, wrote the original draft of the paper, and reviewed and edited the paper. RJ and PAD performed the formal analysis and reviewed and edited the paper. TB, CK, AR, PS, and VP reviewed and edited the paper.

Conflicts of Interest
PS is a recipient of the Robert A Winn Diversity in Clinical Trials Career Development Award, funded by the Bristol Myers Squibb Foundation.

Multimedia Appendix 1
Novel Strategies to Increase Telehealth Engagement (NSITE) Survey.
[PDF File (Adobe PDF File), 1455 KB - formative_v8i1e50572_app1.pdf]

References


Abbreviations

BB: broadband  
EHR: electronic health record  
F2F: face-to-face  
MCHS: Mayo Clinic Health System  
PCP: primary care provider  
SDoH: social determinants of health
provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Transfer of Knowledge on Pneumoconiosis Care Among Rural-Based Members of a Digital Community of Practice: Cross-Sectional Study

Brian Soller\textsuperscript{1}, PhD; Orrin Myers\textsuperscript{2}, PhD; Akshay Sood\textsuperscript{3,4}, MPH, MD

\textsuperscript{1}Department of Sociology, Anthropology, and Public Health, University of Maryland Baltimore County, Baltimore, MD, United States
\textsuperscript{2}Department of Family and Community Medicine, University of New Mexico Health Sciences Center, Albuquerque, NM, United States
\textsuperscript{3}Department of Internal Medicine, University of New Mexico Health Sciences Center, Albuquerque, NM, United States
\textsuperscript{4}Miners Colfax Medical Center, Raton, NM, United States

Corresponding Author:
Akshay Sood, MPH, MD
Department of Internal Medicine
University of New Mexico Health Sciences Center
1 University of New Mexico MSC 10 5550
Albuquerque, NM, 871310001
United States
Phone: 1 5052724751
Fax: 1 5052728700
Email: asood@salud.unm.edu

Abstract

Background: Given the re-emergence of coal workers’ pneumoconiosis in Appalachia and Mountain West United States, there is a tremendous need to train rural professionals in its multidisciplinary management. Since 2016, the Miners’ Wellness TeleECHO (Extension for Community Health Outcomes) Program held by the University of New Mexico, Albuquerque, and Miners’ Colfax Medical Center, Raton, New Mexico, provides structured longitudinal multidisciplinary telementoring to diverse professionals taking care of miners by creating a digital community of practice. Program sessions emphasize active learning through discussion, rather than didactic training. Professional stakeholder groups include respiratory therapists, home health professionals, benefits counselors, lawyers or attorneys, clinicians, and others. Rural-urban differences in knowledge transfer in such a community of practice, however, remain unknown.

Objective: We aim to evaluate the role of the rurality of the patient or client base in the transfer of knowledge to professionals caring for miners using the digital community of practice approach.

Methods: This is a cross-sectional study of 70 professionals participating in the Miners’ Wellness TeleECHO Program between 2018 and 2019. Drawing insights from social network analysis, we examined the association between the rurality of participants’ patient or client base and their self-reported receipt of knowledge. Our focal independent variable was the respondent’s self-reported percentage of patients or clients who reside in rural areas. We measured knowledge transfer sources by asking participants if they received knowledge regarding the care of miners by creating a digital community of practice. Program sessions emphasize active learning through discussion, rather than didactic training. Professional stakeholder groups include respiratory therapists, home health professionals, benefits counselors, lawyers or attorneys, clinicians, and others. Rural-urban differences in knowledge transfer in such a community of practice, however, remain unknown.

Objective: We aim to evaluate the role of the rurality of the patient or client base in the transfer of knowledge to professionals caring for miners using the digital community of practice approach.

Methods: This is a cross-sectional study of 70 professionals participating in the Miners’ Wellness TeleECHO Program between 2018 and 2019. Drawing insights from social network analysis, we examined the association between the rurality of participants’ patient or client base and their self-reported receipt of knowledge. Our focal independent variable was the respondent’s self-reported percentage of patients or clients who reside in rural areas. We measured knowledge transfer sources by asking participants if they received knowledge regarding the care of miners by creating a digital community of practice. Program sessions emphasize active learning through discussion, rather than didactic training. Professional stakeholder groups include respiratory therapists, home health professionals, benefits counselors, lawyers or attorneys, clinicians, and others. Rural-urban differences in knowledge transfer in such a community of practice, however, remain unknown.

Objective: We aim to evaluate the role of the rurality of the patient or client base in the transfer of knowledge to professionals caring for miners using the digital community of practice approach.

Methods: This is a cross-sectional study of 70 professionals participating in the Miners’ Wellness TeleECHO Program between 2018 and 2019. Drawing insights from social network analysis, we examined the association between the rurality of participants’ patient or client base and their self-reported receipt of knowledge. Our focal independent variable was the respondent’s self-reported percentage of patients or clients who reside in rural areas. We measured knowledge transfer sources by asking participants if they received knowledge regarding the care of miners by creating a digital community of practice. Program sessions emphasize active learning through discussion, rather than didactic training. Professional stakeholder groups include respiratory therapists, home health professionals, benefits counselors, lawyers or attorneys, clinicians, and others. Rural-urban differences in knowledge transfer in such a community of practice, however, remain unknown.

Results: Respondents, on average, identified 4.46 (SD 3.16) unique knowledge sources within the community, with a greater number of cross-stakeholder knowledge sources (2.80) than same stakeholder knowledge sources (1.72). The mean knowledge source range was 2.50 (SD 1.29), indicating that, on average, respondents received knowledge sources from roughly half of the 5 stakeholder groups. Finally, the mean heterogeneity of knowledge sources, which can range between 0 and 0.80, was near the midpoint of the scale at 0.44 (SD 0.30). Multivariable analyses revealed that as the rurality of patient or client bases increased, participants reported more knowledge sources overall, more knowledge sources from outside of their stakeholder groups, a higher knowledge source range, and greater heterogeneity of knowledge sources ($P<.05$ for all comparisons).

Conclusions: Our findings suggest that participants who serve rural areas especially benefit from knowledge transfer within the TeleECHO community of practice. Additionally, the knowledge they receive comes from diverse information sources,
emphasizing its multidisciplinary nature. Our results underscore the capacity of the TeleECHO model to leverage technology to promote rural health equity for miners.

*JMIR Form Res 2024;8:e52414*  doi:10.2196/52414

**KEYWORDS**

community of practice; knowledge transfer; pneumoconiosis; telementoring; rural health care; transfer; information; rural; virtual community; lung diseases; lung disease; rural professionals; rural professional; multidisciplinary management; multidisciplinary; miners; miner; health equity

**Introduction**

Recent studies reveal an increasing prevalence and severity of pneumoconiosis (ie, dust-related lung diseases) among US coal workers since the late 1990s [1-7]. Data from the US Coal Workers Health Surveillance Program indicated that the 2017 prevalence of radiographic pneumoconiosis for coal miners with over 25 years of underground mining experience was greater than 10%, which was double the prevalence from the late 1990s. Similarly, the 2014 rate of complicated pneumoconiosis (a particularly deadly form) among long-tenured underground coal miners was 1.1%, compared to 0.3% at its lowest point in the late 1990s [7,8].

This re-emergence of pneumoconiosis presents unique challenges for rural communities. US counties with the highest mortality rates for pneumoconiosis are concentrated in rural contexts with long histories of mining, such as the Rocky Mountain states and central Appalachia [9]. The prevalence of radiographic pneumoconiosis and complicated pneumoconiosis in rural central Appalachian miners, in particular, is much higher than the national average [6]. While the number of miners requiring specialized care has increased, multidisciplinary expertise and access to complex care for pneumoconiosis have decreased in rural areas [10]. Compared with urban residents, residents of rural areas have less access to outpatient pulmonary rehabilitation [11] or pulmonologist services [12]. Rural practitioners also face unique challenges, including professional isolation and complex patient profiles [13], and describe multiple barriers to knowledge acquisition, such as resources and personal costs, physical distance, and time [14]. Such challenges amplify health inequities and mandate innovative approaches to enhance the health and well-being of rural miners, who constitute an underserved, geographically isolated, medically vulnerable, and often underinsured population [10].

Increasing access to education and mentoring of rural professionals involved in the multidisciplinary care of miners can ameliorate the current dearth of skilled expertise in mining-related diseases in rural areas. The multidisciplinary skills required include medicolegal, clinical, and “soft” skills, the latter including the interpersonal and communication skills needed to navigate highly collaborative work in the care of miners. Insufficient expertise among rural providers in these diverse skills demands innovative education and mentoring solutions.

The Miners’ Wellness TeleECHO (Extension for Community Health Outcomes) Program was established in 2016 to provide structured longitudinal multidisciplinary telementoring to members of professional groups caring for miners who reside in pneumoconiosis mortality hotspots in the United States [9]. Professional stakeholder groups include respiratory therapists, home health professionals, benefits counselors, lawyers or attorneys, clinicians, and others. These members from various stakeholder groups constitute a digital community of practice, or a group of people who “share a concern or a passion for something they do and learn how to do better as they interact regularly” [15]. This approach facilitates knowledge transfer and translation among participants. Knowledge transfer refers to the transmission of information and insights between people or groups [16,17], and knowledge translation involves enhancing users’ awareness of multidisciplinary knowledge and its use in day-to-day work and decision-making in the “real world” [16-18]. Importantly, little is known about how digital communities of practice transfer knowledge across professional stakeholder groups that tend to be geographically isolated, such as rural home health care workers and clinicians or specialists. Thus, examining the patterns of knowledge transfer in such communities of practice can provide insight into how technology can be leveraged to enhance care of complex disease in rural settings and how to promote shared objectives within communities of practice.

Preliminary studies indicate a favorable impact of the ECHO telementoring strategies on providers’ self-efficacy in the care of miners [19], adding to the knowledge base about how the ECHO model can enhance the management of other chronic diseases [20-22].

The rurality of the patient bases for those serving miners limits professionals’ capacity to seek and obtain specialized knowledge concerning the care of pneumoconiosis. This specialized knowledge tends to be concentrated within groups from urban areas, where academic health centers are located. Conversely, urban and suburban practitioners may have limited knowledge concerning the day-to-day challenges of rural patients undergoing treatment for complex diseases. The complexity and interdisciplinary nature of care for pneumoconiosis, coupled with the decline of multidisciplinary knowledge sources within rural areas, underscores the need for specialized knowledge transfer to underserved rural areas. Digital communities of practice are well-equipped to transfer multiple kinds of knowledge [23] across stakeholder groups. First, digital meetings help counteract large geographic distances, thereby providing opportunities for transmitting knowledge concerning facts (eg, know-what) and practical knowledge, skills, and expertise (eg, know-how) among otherwise isolated community members. Perhaps equally important, digital communities of practice help members make social connections and leverage their social

https://formative.jmir.org/2024/1/e52414
networks to gain more access to practical skills and best practices and to adapt to the evolving needs of patients [23]. Understanding knowledge transfer between the community of practice participants from urban and rural patient or client bases is, therefore, essential but has not been fully evaluated—in turn, constituting a critical gap in knowledge. Addressing this knowledge gap can inform evidence-based interventions to enhance future efforts aimed at providing interdisciplinary care for rural miners.

This study evaluates the transfer of knowledge to professionals caring for miners using the digital community of practice approach. We integrate methods from social network analysis to examine patterns of knowledge transfer within and across stakeholder groups within a digital community of practice. We consider the association between the rurality of professionals’ patient or client base and (1) the number of knowledge sources from within the community of practice, (2) the number of knowledge sources outside one’s stakeholder group, (3) the number of knowledge sources within one’s stakeholder group, and (4) the range and heterogeneity in knowledge sources across stakeholder groups. Our study represents a crucial step in assessing the potential to reduce health inequity through greater investment in workforce diversity and interprofessional telementoring efforts that promote collaborative health care in medically underserved mining communities. Our study thus has important implications for understanding how technology fused with specialized expertise can be used to address complex health issues within rural, remote, and medically underserved communities and begin to address health inequities rooted in unequal access to medical care, more broadly. This approach may help rural communities counter the re-emergence of the pneumoconiosis epidemic.

**Methods**

**Study Design**

This is a cross-sectional study of professionals participating in the Miners’ Wellness TeleECHO Program, a community-university partnership between a small rural hospital—Miners’ Colfax Medical Center—and its academic partner—University of New Mexico School of Medicine, together constituting the “hub” site of experts. Stakeholder groups include clinical professional groups (clinicians, respiratory therapists, and home health professionals) and nonclinical professional groups (ie, benefits counselors, lawyers or attorneys, and others, including policymakers, administrators, and mine safety officers), constituting the “spoke” sites located across the United States. The hub and spoke partners regularly engage in telementoring and together form a digital community of practice.

**Recruitment**

This study was based on a convenience sample of 70 participants who volunteered to complete this study’s survey, among all program attendees invited, during the 1-year study duration from September 12, 2018, to September 18, 2019. Core program faculty did not participate in the survey.

**Program Description**

As detailed in a previous publication [19], TeleECHO sessions have a uniform format and are held at the same time twice every month, lasting 75 minutes. Program sessions begin with 10-minutes of introductions and announcements, followed by a 15-minute didactic delivered by an invited expert and a 20-minute facilitated question-answer session. Next, the program director facilitates a 30-minute interactive case discussion. Program sessions emphasize active learning through discussion, rather than didactic training. Participants earn continuing medical education (CME) credits without charge, upon completing a CME survey. A multidisciplinary curriculum committee follows a structured curriculum that is continually adapted based on the needs of the learning community, which are identified through review of the CME feedback reports. Attendance at ECHO sessions is open and voluntary, which allows those not presenting a case to view the didactic, partake in case discussions, contribute insight from professional experiences, and learn from the expert panel. Participants can also access experts at hub or spoke sites for urgent consultation outside of program sessions through telephone or electronic correspondence. Recorded didactic sessions are made available through a web-based archive.

**Program Development**

Since July 2016, our program has used the ECHO model to provide long-term and structured telementoring in the care of miners. This approach deviates from traditional telemedicine where providers assume short-term direct care of individual patients [24]. Further, unlike webinars or traditional didactic lectures, the ECHO model provides an interactive discussion of cases with expert panels in real-time that is highly contextualized and adheres to key learning theory principles. As detailed in a previous publication [19], the ECHO model is based on the following five key principles. (1) The model uses internet-based technology for multipoint videoconferencing, to leverage scarce resources. (2) It uses an established disease-management model associated with best evidence for that disease that has been demonstrated to improve outcomes by reducing variation in processes of care and sharing best practices [21,22,25,26]. (3) It uses the principle of case-based learning for participants to learn with guidance from mentors, based on discussion, questions, and investigation of patient cases under their care. Over time, with iterative practice and feedback, participants gain knowledge and skills and progressively become more independent. (4) It creates a digital community of practice, which emphasizes reciprocity in the sharing of skills and information, and acknowledges that all participants bring useful expertise in the care of miners. Through regular interaction, community members increase their own expertise and that of other participants. As a result, the program aims to increase the ability of individual participants to (a) refer miners appropriately to other experts, (b) accept miner referrals from other experts, and (c) to serve as local experts for less experienced community professionals, thereby improving the care of miners. (5) Finally, it uses an internet-based database (ie, iECHO software) to monitor participant outcomes.
Outcomes

We conceptualized knowledge transfer as the transmission of “facts, experiences, and insights” between people or groups [16,17]. We used a social network approach to examine knowledge transfer among community members by measuring respondents’ number of unique knowledge sources. We also considered the stakeholder group where knowledge originates, which allows us to examine the extent to which participants receive interdisciplinary knowledge from others outside of their own stakeholder groups as well as the overall distribution of knowledge sources across stakeholder groups.

We measured knowledge transfer sources by asking participants if they received new and important knowledge regarding the care of miners during and outside of TeleECHO sessions from each of the other participants. To measure knowledge transfer, respondents were given rosters that included names of all registered participants, with the option of providing additional names not on the roster. Rosters were arranged by stakeholder groups to reduce respondent burden and assist recall. We used these nominations to measure our dependent variables that capture unique dimensions of knowledge transfer. Our first dependent variable, number of knowledge sources, is the count of other participants from whom respondents received new and important knowledge (regardless of the source’s stakeholder group).

Apart from the number of knowledge sources, we tested whether rural participants report greater numbers of knowledge sources from outside of their primary stakeholder group. We thus measured the number of cross-stakeholder knowledge sources, which captures the number of participants from whom respondents received knowledge who were outside of respondents’ stakeholder group. We also analyzed the number of same stakeholder knowledge sources, with a measure capturing the number of participants from whom respondents received knowledge that were in the same stakeholder group as the focal respondent.

We also consider 2 dimensions of diversity in the sources of knowledge transfer among respondents. Range captures the extent to which individuals are connected to others from different social systems or interpersonal environments (eg, employers, associations, and schools) [27,28]. Importantly, a higher range level translates to greater access to nonredundant information [29]. We measure knowledge source range by calculating the number of unique stakeholder groups from which respondents reported receiving knowledge. This variable ranges from 0 to 5, with 0 indicating respondents reported no knowledge sources to 5 indicating respondents received knowledge from at least one member from each of the 5-stakeholder groups.

Our second measure capturing the diversity of knowledge sources is heterogeneity of knowledge sources. Our measure of heterogeneity of knowledge sources taps the distribution of knowledge sources across stakeholder groups for each respondent and is calculated as follows [30]:

Here, \( A_j \) is the number of knowledge sources that belong to a stakeholder group \( j \), \( k_s \) is the number of knowledge sources, and \( n \) is the total number of stakeholder groups from which the focal respondent reported receiving knowledge (ie, knowledge source range). Heterogeneity increases when respondents receive knowledge from a larger number of different stakeholder groups (ie, have high knowledge source range) and the knowledge sources are equally distributed across the stakeholder groups. In our study, this measure potentially ranges from 0 to 0.8, with higher values indicating greater heterogeneity in knowledge sources. Note, heterogeneity is undefined for respondents reporting 0 knowledge sources, which was the case for 2 of our respondents, who were excluded for analyses of heterogeneity.

Independent Variables

Our focal independent variable captures the level of rurality among a respondent’s patient or client base. The measure is based on the percent of patients or clients who reside in rural areas, as self-reported by the participant. Initial responses were ordinal and included five categories: 1 (0% to 20%), 2 (21% to 40%), 3 (41% to 60%), 4 (61% to 80%), and 5 (81% to 100%). For this study, we collapsed the ordinal variable into a binary variable indicating rural patient or client base, which equals 1 if 41% to 100% of a respondent’s patient or client base resided in rural areas, and 0 if only 0% to 40% of their patients or clients lived in rural areas. We collapsed the categories for 2 reasons. First, exploratory analyses revealed that only 5 respondents reported serving a 21% to 40% rural patient or client base. Second, comparisons of the means of the knowledge source variables across levels of patient or client rurality suggested a threshold effect, with only minimal differences in the outcomes for those serving 0% to 20% versus 21% to 40% rural patient or client base but large differences between these combined categories and those serving a 41% or greater rural patient or client base. The results based on the original 5-category ordinal variable were similar to those presented in this study.

Covariates

Multivariable models include control variables to account for potential confounding between the association between patient or client rurality and our outcomes. Experienced versus fresh participant: fresh participants were defined as those who first attended the community of practice in or after the summer of 2018 (defined as from May 9, 2018, onwards) versus experienced participants (defined by those who had first attended any time between July 1, 2016, and May 8, 2018). Experienced participants had greater cumulative participation and therefore, experience with the TeleECHO Program than fresh participants (11.4, SD 9.8 vs 4.6, SD 4.6 total sessions attended before or during this study’s timeframe; \( P<.03 \)). This cutoff date was chosen based on the date of funding by the sponsor, which allowed the frequency of the TeleECHO Program to be raised from monthly to twice a month. Respondents’ length of care for miners taps the number of years each participant reported having served miners. Initial responses were measured in years. To aid in the interpretation of our regression results, we divide the reported number of years cared for miners by 10, so that the variable captures the number of decades respondents reported having cared for miners.
We also control for participant demographics. We control for age with 2 dummy variables indicating 51 to 60 years old and older than 60 years (1=yes, 0=no) with younger than 50 years old serving as the reference category. Male sex is binary and indicates the respondent reported a male sex identity (1=yes, 0=no). Respondents reported their race and Hispanic ethnicity status. Based on the responses from these questions, respondents were initially categorized as either Asian, non-Hispanic-Black or African American, Hispanic, multiracial or some other race, or non-Hispanic White. We report the percentages of respondents in each race or ethnic category but collapsed categories into a binary variable indicating non-White (1=yes, non-Hispanic White is the reference) in our regression analyses due to the small sizes of non-White racial or ethnic groups in the sample. Alternative methods of collapsing race or ethnic categories resulted in similar findings as those presented here.

Data Collection
The program monitored the number of sessions, learners, unique learners, geographical sites of learners, and patient cases presented (using the iECHO software). Survey data were collected using the REDCap (Research Electronic Data Capture; Vanderbilt University), a secure web app for building and managing online surveys and databases.

Analytic Strategy
All analyses were conducted in Stata/MP (version 16.0; StataCorp LLC). We used negative binomial regressions to analyze the total number of knowledge sources, number of cross-stakeholder knowledge sources, and number of same stakeholder knowledge sources, which were discrete counts and were over dispersed. Ordinary least squares regression was used to analyze knowledge source diversity and knowledge source heterogeneity. Model coefficients (b) and SE were used to summarize effect sizes. Data missingness due to nonresponse was minimal, with 2 respondents declining to report their age, 1 respondent declining to report on length of care for miners, and 1 respondent declining to report on rurality (for a total of 3 respondents having missing data on at least 1 variable). Missing values on these measures were imputed using the Stata ice procedure [31], and models were estimated with 10 imputed data sets using the mi command suite in Stata 16. The results based on unimputed data using listwise deletion were nearly identical to those presented here.

Ethics Considerations
Approval was obtained from the institutional review board, Human Research Protections Office, at the University of New Mexico Health Sciences Center (18-386). Anonymized consent was obtained from all participants. Study data were deidentified for analysis to maintain confidentiality. All participants were provided an electronic merchandise card of US $50 upon survey completion.

Results
Table 1 shows the descriptive characteristics of the 70 ECHO participants caring for pneumoconiosis in a cross-sectional study during the timeframe of 2018-2019.
Table 1. Descriptive characteristics of study participants (N=70).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Knowledge source variables, mean (SD)</strong></td>
<td></td>
</tr>
<tr>
<td>Number of knowledge sources (N=70)</td>
<td>4.46 (3.16)</td>
</tr>
<tr>
<td>Same stakeholder knowledge sources (n=61)</td>
<td>1.72 (1.46)</td>
</tr>
<tr>
<td>Cross-stakeholder knowledge sources (n=61)</td>
<td>2.80 (2.63)</td>
</tr>
<tr>
<td>Knowledge source range (N=70)</td>
<td>2.50 (1.29)</td>
</tr>
<tr>
<td>Heterogeneity of knowledge sources (n=68)</td>
<td>0.44 (0.30)</td>
</tr>
</tbody>
</table>

Age group (y), n (%)  
- 50 or younger: 36 (51)  
- 51 to 60: 15 (21)  
- Older than 60: 19 (27)  

Sex, n (%)  
- Female: 45 (64)  
- Male: 25 (36)  

Race or ethnicity, n (%)  
- Asian: 6 (9)  
- Hispanic: 5 (7)  
- Non-Hispanic Black or African American: 1 (1)  
- Non-Hispanic White: 55 (79)  
- Other: 3 (4)  

Primary stakeholder group, n (%)  
- Clinician: 20 (29)  
- Respiratory therapist: 12 (17)  
- Lawyer or attorney: 7 (10)  
- Benefits counselor: 8 (11)  
- Home health professional: 14 (20)  
- Others: 9 (13)  

Rurality of patient or client base, n (%)  
- Nonrural patient or client base: 19 (27)  
- Rural patient or client base: 51 (73)  

Participant experience, n (%)  
- Fresh participant: 40 (57)  
- Experienced participant: 30 (43)  

Decades serving miners (N=70), mean (SD): 0.76 (0.72)  

Knowledge Source Variables Among all Participants  
Respondents, on average, identified 4.46 (SD 3.16) unique knowledge sources within the community. Respondents, on average, reported greater numbers of cross-stakeholder knowledge sources (2.80, SD 2.63) than same stakeholder knowledge sources (1.72, SD 1.46). The mean knowledge source range was 2.50 (SD 1.29), indicating that, on average, respondents received knowledge sources from roughly half of the 5 stakeholder groups. Finally, the mean heterogeneity of knowledge sources, which can range between 0 and 0.80, was near the midpoint of the scale at 0.44 (SD 0.30).  

Knowledge Source Variables Among Participants Serving Rural Versus Nonrural Bases  
We explain the means of the knowledge source measures among those serving rural versus nonrural patient or client bases. Those serving rural patient or client bases, on average, reported 5.00 (SD 3.13) unique knowledge sources compared to 3.00 (SD 2.88) among those primarily serving nonrural patients or clients. There was only a minor difference in the mean number of same
stakeholder knowledge sources for those serving rural (1.68, SD 1.42) versus nonrural (1.86, SD 1.66) patients or clients. However, rural providers, on average, identified 3.30 (SD 2.61) cross-stakeholder knowledge sources, whereas nonrural providers, on average, identified 1.14 (SD 1.99) cross-stakeholder knowledge sources. Finally, comparing the measures of diversity of knowledge sources, those serving rural patients or clients had a higher mean knowledge source range 2.78 (SD 1.22) versus 1.74 (SD 1.19) and mean knowledge source heterogeneity 0.52 (SD 0.25) versus 0.23 (SD 0.31) than those serving primarily nonrural patients or clients.

**Multivariable Results**

Table 2 presents results from multivariable regression models of the different dimensions of knowledge transfer.

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Model 1: number of knowledge sources (N=70)</th>
<th>Model 2: same stakeholder knowledge sources (n=61)</th>
<th>Model 3: cross-stakeholder knowledge sources (n=61)</th>
<th>Model 4: knowledge source range (n=70)</th>
<th>Model 5: heterogeneity of knowledge, sources (n=68)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b (SE)</td>
<td>P value</td>
<td>b (SE)</td>
<td>P value</td>
<td>b (SE)</td>
</tr>
<tr>
<td>Rurality</td>
<td>0.50 (0.22)</td>
<td>.02</td>
<td>0.04 (0.26)</td>
<td>.89</td>
<td>0.91 (0.37)</td>
</tr>
<tr>
<td>Age (y; ≤50 y is the reference)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51-60</td>
<td>-0.05 (0.24)</td>
<td>.84</td>
<td>0.67 (0.30)</td>
<td>.03</td>
<td>-0.33 (0.34)</td>
</tr>
<tr>
<td>&gt;60</td>
<td>-0.35 (0.28)</td>
<td>.22</td>
<td>0.14 (0.34)</td>
<td>.68</td>
<td>-0.52 (0.40)</td>
</tr>
<tr>
<td>Male sex (female sex is the reference)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-White (non-Hispanic White is the reference)</td>
<td>0.32 (0.22)</td>
<td>.14</td>
<td>-0.24 (0.30)</td>
<td>.41</td>
<td>0.62 (0.34)</td>
</tr>
<tr>
<td>Stakeholder group (clinical provider is the reference)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory therapist</td>
<td>0.15 (0.27)</td>
<td>.58</td>
<td>-0.50 (0.36)</td>
<td>.17</td>
<td>0.54 (0.37)</td>
</tr>
<tr>
<td>Lawyer or attorney</td>
<td>-0.04 (0.32)</td>
<td>.89</td>
<td>-0.50 (0.41)</td>
<td>.23</td>
<td>0.36 (0.45)</td>
</tr>
<tr>
<td>Benefits counselor</td>
<td>0.29 (0.29)</td>
<td>.31</td>
<td>-0.82 (0.43)</td>
<td>.06</td>
<td>0.84 (0.39)</td>
</tr>
<tr>
<td>Home health professional</td>
<td>0.34 (0.25)</td>
<td>.18</td>
<td>0.20 (0.30)</td>
<td>.50</td>
<td>0.38 (0.36)</td>
</tr>
<tr>
<td>Others</td>
<td>0.20 (0.29)</td>
<td>.50</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Experienced participant (fresh participant is the reference)</td>
<td>0.37 (0.17)</td>
<td>.03</td>
<td>-0.17 (0.21)</td>
<td>.42</td>
<td>0.74 (0.25)</td>
</tr>
<tr>
<td>Decades serving miners</td>
<td>0.18 (0.16)</td>
<td>.24</td>
<td>0.08 (0.19)</td>
<td>.68</td>
<td>0.17 (0.24)</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.68 (0.30)</td>
<td>.02</td>
<td>0.66 (0.35)</td>
<td>.06</td>
<td>-0.58 (0.45)</td>
</tr>
</tbody>
</table>

**Number of Knowledge Sources**

Model 1 examines the number of knowledge sources using a negative binomial regression. The results indicate that rural patient or client base is positively associated with the number of knowledge sources (b=0.50; P=.02). This finding suggests that providers serving rural clients or patients identify greater numbers of knowledge sources within the community of practice than participants whose patients or clients reside in nonrural areas, even after accounting for key confounders. Compared with fresh participants, experienced participants report greater numbers of knowledge sources (b=0.37; P=.03). No other participant characteristics were significantly associated with the number of knowledge sources (all P>.05).

**Number of Same Stakeholder and Cross-Stakeholder Knowledge Sources**

Models 2 and 3 examine the number of same stakeholder and cross-stakeholder knowledge sources, respectively. These models provide insight into whether participants tend to identify...
knowledge sources from within or outside of their primary stakeholder groups. Note, participants from “Other” stakeholder groups were dropped from Models 2 and 3, as they by definition have all different stakeholder ties and 0 same stakeholder ties. The results from Model 2 indicate that rural patient or client base has a nonsignificant association with the number of same stakeholder knowledge sources ($b=0.04; P=.89$). Compared with those aged 50 years or younger, participants between the ages of 51 and 60 years report more ties to members of the same stakeholder group ($b=0.67; P=.03$). Turning to Model 3, which examines the number of cross-stakeholder knowledge sources, theurrality of the patient or client base was positively associated with the number of cross-stakeholder ties ($b=0.91; P=.01$). This indicates that participants serving larger proportions of rural patients or clients reported larger numbers of cross-stakeholder knowledge sources than those serving smaller proportions of rural patients or clients. Additionally, experienced participants report larger numbers of cross-stakeholder knowledge sources than fresh participants ($b=0.74; P=.003$) and benefits counselors report larger numbers of cross-stakeholder knowledge sources than clinicians ($b=0.84; P=.03$).

**Range and Heterogeneity of Knowledge Sources**

The final models in Table 2 examine the range and heterogeneity of participants’ knowledge sources. These models provide insight into the number of different stakeholder groups from which participants received knowledge, and the extent to which participants’ knowledge sources are equally dispersed across different stakeholder groups. Turning to Model 4, which is a linear regression of knowledge source range, we found that a rural patient or client base has a positive and significant coefficient ($b=0.92; P=.01$). Model 4 also indicates that experienced participants reported higher knowledge source range than fresh and new participants ($b=0.70; P=.03$).

Finally, Model 5 examines the association between rurality and participants’ knowledge source heterogeneity. Whereas range is the count of the number of unique stakeholder groups from which participants receive knowledge, knowledge source heterogeneity also assesses whether stakeholders from which participants receive knowledge tend to be concentrated in 1 stakeholder group (low heterogeneity) versus equally distributed across multiple groups (high heterogeneity). Note, that because knowledge source heterogeneity can only be measured among participants with at least one knowledge source, Model 5 excludes 2 respondents who reported 0 knowledge sources. Patient or client rurality was positively associated with participants’ heterogeneity of knowledge sources ($b=0.25; P=.003$), indicating knowledge sources are more equally distributed across stakeholder groups as the rurality of their patient and client bases increased.

**Discussion**

**Principal Findings**

Community of practice participants with higher proportions of rural patient or client base, on average, report more knowledge sources overall, more knowledge sources from outside of their stakeholder groups, a higher knowledge source range, and greater heterogeneity of knowledge sources than those with a lower proportion of rural patient or client base. These findings were confirmed after adjustment for potential confounders in regression analyses. More broadly, these findings suggest participants who serve rural areas especially benefit from knowledge transfer within the TeleECHO community of practice. Additionally, the knowledge they receive comes from diverse information sources, emphasizing its multidisciplinary nature.

Further, 1 primary objective of Project ECHO is to decentralize knowledge for the care of patients through exchanging insights and information. Knowledge transfer is key to enhancing the care of complex disease by timely, evidence-based information shared by experts who have used, amplified, and applied this knowledge with interested professionals who (1) are seeking knowledge to assist their patients or clients and (2) through its application, increase access to complex disease care for patients in rural and underserved communities. Project ECHO supports knowledge transfer within the community of practice, through experts sharing and discussing evidence in association with challenging questions with which professionals at program spoke sites are wrestling. Our study suggests this knowledge transfer may be particularly effective among professionals with longer experience with the program.

Professionals in rural mining communities often lack access to traditional knowledge sources. This disparity results from professional isolation; challenges with continuing professional education that requires travel to a distant site for participation with resultant closure of their practices, often without adequate coverage available; and unavailability of specialists with more in-depth knowledge about the clinical, medicolegal, and interpersonal aspects of care of miners. The need to increase access to information for rural professionals is, therefore, obvious. To this end, information technology has come to the fore. However, research suggests that even when electronic information services are provided to rural practitioners, they may not be well used [32]. The lack of information handling skills, lack of time, and perceived peripherality to the job are all seen as major constraints [33,34]. However, our study challenges this belief by demonstrating that professionals serving rural areas especially benefit from access to knowledge through the innovative TeleECHO model, which would otherwise remain siloed within stakeholder groups. Further, the knowledge source range and heterogeneity that the TeleECHO model promotes may allow greater access to thought-provoking ideas that foster learning and other growth-enhancing actions [27,35]. To the best of our knowledge, our approach of studying patterns of knowledge transfer, using social network analysis tools, has never been used previously.

**Strengths**

Our study has multiple strengths. It involves an innovative intervention that addresses the barriers to the care of miners by using the TeleECHO model, which provides a multidisciplinary community of practice approach, using internet-based technology, an approach that has been well studied in other diseases [21,22,25,26]. This study is topical and significant because it addresses a critical gap related to the emerging pneumoconiosis epidemic in the rural United States. Since the
ECHO model has been adopted nationally and globally to improve rural access to care for patients with numerous chronic diseases, there already exists infrastructure to allow for rapid scaling of the Miners’ Wellness TeleECHO Program nationally and globally.

Limitations
There are also limitations to this study. We are unable to correlate knowledge transfer to patient outcomes or changes in provider behavior. We have, however, previously published a listing of qualitative changes that our ECHO participants reported they were going to make in their practice, obtained as part of a CME survey requested at the end of each TeleECHO session [36]. Although a small sample size raises the possibility of a type I error, individual professionals and teams of professionals trained in the ECHO model can reach a large number of miners, with the potential for creating exponential change. High-risk individuals who did not volunteer to participate in this study would not have provided information in the estimation of the program effects, thus introducing an element of potential participation bias. The knowledge transfer instrument was not validated in this study. Program participants had variable competencies, with varying levels of sophistication, commitment, expertise, experience, and historic levels of collaboration within the TeleECHO Program. However, adjustment for participant experience with the TeleECHO Program or length of care for miners in the multivariable models did not change our study findings. Intergenerational, interinstitutional, and rural-urban disparities in ability to leverage technology by participating professionals may challenge empirical examinations of knowledge transfer. Finally, data limitations, including survey nonparticipation by the core program faculty and survey nonresponse among the TeleECHO Program participants, preclude the use of complex social network analysis methods (eg, exponential random graph models) commonly used to examine network selection processes in our study. Although our methods are adequate for examining associations between participating characteristics and the number, range, and heterogeneity of knowledge sources, we are unable to examine how network processes such as reciprocal knowledge transfer operate within the learning community.

Conclusions
Despite these limitations, our findings suggest participants who serve rural areas especially benefit from knowledge transfer within the TeleECHO community of practice. Additionally, the knowledge they receive comes from diverse information sources, emphasizing its multidisciplinary nature. Our results underscore the capacity of the Project ECHO model to leverage technology and workforce diversity to facilitate knowledge transfer to rural professionals and ultimately promote health equity among rural and medically underserved mining communities. Although this approach addresses a critical gap related to the emerging pneumoconiosis epidemic in rural United States, future research will evaluate whether this translates into improved patient outcomes in rural mining communities.

Acknowledgments
This study was funded by the Alpha Foundation (grant AFC719-14). The sponsors played no role in the conduct, analysis, and dissemination of this study and in the review and approval of this paper for publication.

Data Availability
The data sets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions
All authors were involved with the following: (1) substantial contributions to the conception or design of the work (BS, OM, and AS) or the acquisition (BS, OM, and AS), analysis (BS and OM), or interpretation of data for the work (BS, OM, and AS); (2) drafting the work or reviewing it critically for important intellectual content (BS, OM, and AS); (3) final approval of the version to be published (BS, OM, and AS); and (4) agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved (BS, OM, and AS).

Conflicts of Interest
None declared.

References


Abbreviations

CME: continuing medical education
ECHO: Extension for Community Health Outcomes
REDCap: Research Electronic Data Capture

©Brian Soller, Orrin Myers, Akshay Sood. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Medical Students’ Perceptions on Identifying and Addressing Emotional Responses in Emergency Medicine: Pilot Investigation

Anish Kumar Agarwal¹,², MD, MPH, MS; Rachel Gonzales¹,², MPH; Cory Munden¹, MD; DaCarla Albright³, MD; Suzana Tsao¹, DO

¹Department of Emergency Medicine, University of Pennsylvania, Philadelphia, PA, United States
²Center for Health Care Transformation and Innovation, Penn Medicine, Philadelphia, PA, United States
³Department of Obstetrics and Gynecology, University of Pennsylvania, Philadelphia, PA, United States

Abstract

Background: Training in acute care, such as emergency medicine (EM), where exposure to critically ill and injured patients is high, impacts the well-being of trainees and contributes to burnout. Investigating how, and if, trainees prepare for these situations is necessary to ensure they are supported.

Objective: This study aimed to evaluate medical students’ perspectives and emotional preparedness for handling acute care and trauma.

Methods: We conducted a pilot investigation using a remote digital survey of medical students during their EM clerkship at a large, urban academic institution. The primary outcome of interest was student-reported preparedness and comfort in handling trauma and critical care patient encounters. Secondary outcomes included awareness of well-being resources and comfort in accessing digital well-being resources.

Results: A total of 57 medical students completed the voluntary digital survey, and half of the students (n=28, 49%) reported having witnessed the care of a critically ill or a penetrating trauma patient (eg, a victim of gun violence). A majority (n=40, 70%) had thought about how these events may impact them, and over half felt unprepared to identify the emotional impact these cases may have on them (n=31, 54%) or address the emotional or mental health impact (n=36, 63%). Less than a quarter (n=14, 25%) were aware of digital mental health resources, and 58% (n=33) did not feel fully comfortable connecting with resources if needed. Students who had previously witnessed critical care were significantly more likely to report feeling well prepared in identifying the emotional impact and addressing this impact.

Conclusions: In this cross-sectional survey, students did not feel fully prepared to identify or address the emotional impact of working in EM. Additionally, they lacked awareness of or comfort with accessing digital institutional resources meant to support their well-being, such as a large web-based platform. These findings can help inform and guide interventions by educational and academic leaders. The aim would be to create and promote environments that empower students with tools to identify their own emotions and connect to well-being resources.

(JMIR Form Res 2024;8:e50827) doi:10.2196/50827

KEYWORDS

well-being; burnout; medical education; coping; student; students; university; college; acute care; trauma; traumatic; emotion; emotional; stress; distress; psychological; cross-sectional; survey; surveys; critical; critically; perception; perspectives; prepared; preparedness
Introduction

Health care–associated burnout persists in medicine and can be identified early in medical training [1-4]. This syndrome has negative impacts at the individual level, and it also affects patient care and the health care system by contributing to more medical errors, lower patient satisfaction, reduced productivity, poor clinical teaching and role modeling for trainees, increased cost, and increased physician attrition and ultimately contributes to physician suicide [5-8]. Physicians have a higher risk for suicide than the general population, and these mental health risks have been labeled as known “occupational hazards” [9,10]. Notably, the rate of US physicians annually completing suicide is estimated to be equivalent to the number of students in 3 graduating medical school classes [11,12].

Studies have demonstrated that the impacts of burnout and mental health symptoms, such as depression and anxiety, begin early in medical education and are found in medical students, residents, and physicians in training [2,13-15]. Studies have also found that rates of depression are higher for medical students than other trainees, suggesting this group may be particularly in need of interventions to support well-being and prevent burnout [11,14]. The strain and emotional toll of working in health care emerges early in training. For students, investigating how one identifies, processes, and copes with feelings of anxiety, sadness, depression, or stress related to their clinical experiences within clerkships remains understudied [16]. Emergency medicine (EM) physicians consistently report some of the highest rates of burnout, with EM often being referred to as the “center” of burnout [6,17]. EM physicians are 3 times more likely to be burned out compared to non-EM physicians [4].

Investing in strategies to help trainees identify their own emotions related to providing care, how to cope, and how to sustain well-being is critical for the future of the workforce [18]. Rather than awaiting burnout to evolve, examining how prepared these students feel is necessary to ensure that health systems and medical schools adequately support and proactively maintain medical students’ well-being. This is especially important given the rise of mental health symptoms and burnout in health care within the backdrop of the COVID-19 pandemic and social unrest related to racial injustice and rising gun violence.

The goal of this study was to assess medical students’ preparation during their EM rotation to understand how students self-identify their capacity to deal with emotionally charged clinical settings (eg, critical care cases or trauma).

Methods

Ethical Considerations

This pilot investigation used an electronic voluntary survey administered to second- and third-year medical students at the University of Pennsylvania in Philadelphia during their EM clerkship at a large, urban academic institution. The study was approved by the University of Pennsylvania Institutional Review Board (849318). All research methods, consent, and activities were performed in accordance with the university guidelines and regulations.

Eligible Participants and Study Type

Inclusion criteria consisted of students in their second (preclerkship or M2) or third (clerkship or M3) year of medical school. This study was cross-sectional, and a voluntary response sample was used. There were no specific exclusion criteria, as other students were not invited to the survey.

Recruitment Procedures

Participants were invited via email, completed informed consent, and were not compensated for participating. Data were collected and aggregated for analysis. The students at this medical school are routinely surveyed, and questions from this study were incorporated into the preexisting and ongoing school surveying. In total, the survey was sent to 161 students in their preclerkship (M2) and 153 students in their clerkship (M3) during the final week of classes in December 2021.

Approach and Analysis

This cross-sectional pilot study was developed by the research team, with expertise in medical education (authors ST and DA), qualitative methods (AKA and RG), and clinician well-being (AKA). No previous instrument, to the knowledge of the study team, exists; thus, the instrument was developed and pilot-tested in this study. All answers were anonymous, and no demographic information was collected. The primary outcome of interest was student-reported preparedness and comfort in handling trauma and critical care patient encounters. Secondary outcomes included awareness of well-being resources available to them during their clerkship, feelings of preparedness, and comfort in accessing well-being resources. Comparisons were done using chi-square tests in Stata IC 16.1 (StataCorp), and a P<.05 was considered statistically significant.

Results

A total of 57 medical students completed the voluntary survey; 26 (46%) of them were M2 students, and 31 (54%) were M3 students. Almost half (n=28, 49%) of the students reported having witnessed the care of a critically ill or injured patient (defined as a victim of gun violence). Most (n=40, 70%) students had thought about how these events may impact them, but most did not feel fully prepared to identify the emotional impact these cases may have on them (n=31, 55%) or prepared to address this emotional or mental health impact (n=36, 63%). Although resources are widely available to support students’ well-being at this institution, only 25% (n=14) were aware of these institutional resources to help them cope with the emotions involved in care, and 58% (n=33) did not feel fully comfortable connecting with resources if needed (Table 1).

Differences were identified between those students who had witnessed the care of a critically ill or injured patient and those who had not (Table 2). Students who had witnessed such care were more likely to feel well prepared in identifying the emotional impact of these cases (n=7, 25% vs n=0, 0%; P=.007) and in addressing this impact (n=9, 32% vs n=0, 0%; P=.001). No significant differences were found in student awareness of...
resources and their comfort in connecting with these resources to cope with the emotions involved with care.

Table 1. Medical students’ perspectives on the emotional impact and preparedness of caring for critical patients.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Values (N=57)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you witnessed the immediate care of a critically ill or injured patient?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>28 (49)</td>
</tr>
<tr>
<td>No</td>
<td>29 (51)</td>
</tr>
<tr>
<td>Have you thought about how these events may impact you?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>40 (70)</td>
</tr>
<tr>
<td>No</td>
<td>15 (26)</td>
</tr>
<tr>
<td>N/A</td>
<td>2 (4)</td>
</tr>
<tr>
<td>How prepared do you feel in identifying the emotional impact these cases may have?</td>
<td></td>
</tr>
<tr>
<td>Well prepared</td>
<td>7 (12)</td>
</tr>
<tr>
<td>Prepared</td>
<td>18 (32)</td>
</tr>
<tr>
<td>Somewhat prepared</td>
<td>23 (40)</td>
</tr>
<tr>
<td>Not prepared</td>
<td>8 (14)</td>
</tr>
<tr>
<td>N/A</td>
<td>1 (2)</td>
</tr>
<tr>
<td>How prepared do you feel in addressing with the emotional impact these cases may have?</td>
<td></td>
</tr>
<tr>
<td>Well prepared</td>
<td>9 (16)</td>
</tr>
<tr>
<td>Prepared</td>
<td>11 (19)</td>
</tr>
<tr>
<td>Somewhat prepared</td>
<td>25 (44)</td>
</tr>
<tr>
<td>Not prepared</td>
<td>11 (19)</td>
</tr>
<tr>
<td>N/A</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Are you aware of university resources to support learners as they cope with the emotions involved with care?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>14 (25)</td>
</tr>
<tr>
<td>No</td>
<td>42 (74)</td>
</tr>
<tr>
<td>N/A</td>
<td>1 (2)</td>
</tr>
<tr>
<td>How comfortable you would feel in connecting with resources if needed?</td>
<td></td>
</tr>
<tr>
<td>Very comfortable</td>
<td>7 (12)</td>
</tr>
<tr>
<td>Comfortable</td>
<td>16 (28)</td>
</tr>
<tr>
<td>Somewhat comfortable</td>
<td>23 (40)</td>
</tr>
<tr>
<td>Not comfortable</td>
<td>10 (18)</td>
</tr>
<tr>
<td>N/A</td>
<td>1 (1.8)</td>
</tr>
</tbody>
</table>

aN/A: not applicable.
Table 2. Medical students’ perspectives on personal emotional reaction, preparation, and coping skills based on prior exposure.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Ever witnessed the care of a critically ill or injured patient (N=57)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes (n=28)</td>
</tr>
<tr>
<td>Have you thought about how these events may impact you?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>22 (78.6)</td>
</tr>
<tr>
<td>No</td>
<td>6 (21.4)</td>
</tr>
<tr>
<td>N/A*</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>How prepared do you feel in identifying the emotional impact these cases may have?</td>
<td></td>
</tr>
<tr>
<td>Well prepared</td>
<td>7 (25.0)</td>
</tr>
<tr>
<td>Prepared</td>
<td>10 (35.7)</td>
</tr>
<tr>
<td>Somewhat prepared</td>
<td>10 (35.7)</td>
</tr>
<tr>
<td>Not prepared</td>
<td>1 (3.6)</td>
</tr>
<tr>
<td>N/A*</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>How prepared do you feel in addressing with the emotional impact these cases may have?</td>
<td></td>
</tr>
<tr>
<td>Well prepared</td>
<td>9 (32.1)</td>
</tr>
<tr>
<td>Prepared</td>
<td>6 (21.4)</td>
</tr>
<tr>
<td>Somewhat prepared</td>
<td>12 (42.9)</td>
</tr>
<tr>
<td>Not prepared</td>
<td>1 (3.6)</td>
</tr>
<tr>
<td>N/A*</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Are you aware of university resources to support learners as they cope with the emotions involved with care?</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>9 (32.1)</td>
</tr>
<tr>
<td>No</td>
<td>19 (67.9)</td>
</tr>
<tr>
<td>N/A*</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>How comfortable you would feel in connecting with resources if needed?</td>
<td></td>
</tr>
<tr>
<td>Very comfortable</td>
<td>5 (17.9)</td>
</tr>
<tr>
<td>Comfortable</td>
<td>8 (28.5)</td>
</tr>
<tr>
<td>Somewhat comfortable</td>
<td>11 (39.3)</td>
</tr>
<tr>
<td>Not comfortable</td>
<td>4 (14.3)</td>
</tr>
<tr>
<td>N/A*</td>
<td>0 (0.0)</td>
</tr>
</tbody>
</table>

*aN/A: not applicable.

Discussion

Principal Findings

Given EM’s high rates of burnout and the vulnerable role medical students hold as trainees, medical students undergoing their EM clerkship are at high risk of emotional strain and stress [15]. The rising focus on physician mental health extends to those in training [8]. There is a gap in understanding how prepared medical students feel in identifying and addressing their emotional response and the impact experiences in EM may have upon them. This study assessed medical students’ preparation to better understand their needs and guide interventions toward key priority areas of focus.

This study found that, regardless of whether students had already witnessed the care of a critically ill patient, most students did not feel prepared to identify or address the emotional impact associated with these situations. This highlights the need to train students early and normalize the emotional impact of working in medicine [19,20]. The stigma associated with mental health in health care has pervaded the classroom and hospitals, and to build structures to prevent burnout, we must begin early to help trainees identify feelings of anxiety, depression, or stress as they experience them [13,21-23]. A proactive approach would provide students with the tools and resources they need to adequately identify their emotions and connect to appropriate resources when needed.

It is essential for institutions not only to have resources available to support students’ well-being and mental health but also to make these resources readily accessible and easy to navigate. Our findings reveal that even in an environment where resources are present, students may be unaware of these resources or may not be comfortable accessing them. The University of Pennsylvania School of Medicine has a robust infrastructure within the medical school, a separate web-based mental health
and well-being platform, accessible to the entire health system community [24]. However, students in this study remained unaware of their availability or accessibility. Institutions must work to incorporate these resources into the clinical and teaching environments to reduce the stigma that may prevent students from accessing them.

Finally, the significant differences in feelings of preparedness between those students who had and those who had not witnessed the care of critical care patients suggest that students do not feel prepared to identify the emotional impact such experience may have until it has happened. It is important to act proactively to prepare students to experience the care of critical patients. EM provides an ideal environment to do so, as the likelihood that students will be placed in an emotionally charged setting is high [17,25]. Venues such as critical care, pediatrics, obstetrics and gynecology, as well as surgery provide other opportune areas for schools to think about deploying focused interventions to where students may need them most.

Limitations
This study has some limitations. First, to protect confidentiality and privacy, we did not collect demographic information. This prevented us from analyzing how these findings might differ by age, race, or ethnicity. Second, as medical students self-selected to participate, selection bias may play a role. The participants surveyed here may not accurately represent the experiences of all medical students in the clerkship program. Additionally, this study was performed at a single urban, academic program and may not be applicable to all medical students at other various institutions. We are also limited by a sample of 56 students, which may further limit our ability to generalize these findings.

Conclusions
Similar to other roles in health care, medical students do not feel fully prepared to identify or address the emotional impact of working in acute care. Additionally, they lack awareness of or comfort with accessing institutional resources designed to support their well-being. Medical students who have not witnessed the care of a critically ill or injured patient were more likely to feel unprepared in identifying or addressing the emotional impact such an event might have on them. These findings can help inform and guide interventions by educational and academic leaders. The aim would be to create and promote training environments that empower students with tools to identify their own emotions and connect to well-being resources. We need to normalize the conversation around mental health in the health care workforce and reduce stigma early in medicine, beginning with our medical students.

Acknowledgments
We would like to acknowledge the students and staff of the University of Pennsylvania Department of Emergency Medicine.

Data Availability
The data sets generated and analyzed during the study are not publicly available due to the sensitivity of data content in responses, but deidentified data are available by email request to the corresponding author upon reasonable request.

Authors' Contributions
AKA planned, executed, and led this study. RG planned data collection and completed analysis. CM drafted and revised the manuscript. DA executed the study and revised the manuscript. ST planned and supervised the study. All authors contributed to the draft, revisions, and manuscript preparation.

Conflicts of Interest
None declared.

References
Abbreviations

EM: emergency medicine
Machine Learning for Early Prediction of Major Adverse Cardiovascular Events After First Percutaneous Coronary Intervention in Patients With Acute Myocardial Infarction: Retrospective Cohort Study

Pin Zhang¹,²*, MM; Lei Wu¹*, MM; Ting-Ting Zou¹, MM; ZiXuan Zou¹, MM; JiaXin Tu¹, MM; Ren Gong³*, PhD; Jie Kuang¹*, PhD

¹Jiangxi Provincial Key Laboratory of Preventive Medicine, School of Public Health, Nanchang University, Nanchang, China
²School of Public Health and Management, Nanchang Medical College, Nanchang, China
³Department of Cardiology, The Second Affiliated Hospital of Nanchang University, Nanchang, China
*these authors contributed equally

Corresponding Author:
Jie Kuang, PhD
Jiangxi Provincial Key Laboratory of Preventive Medicine
School of Public Health
Nanchang University
461 BaYi Road
Nanchang, 330006
China
Phone: 86 0791 8636228
Email: kuangjie@ncu.edu.cn

Abstract

Background: The incidence of major adverse cardiovascular events (MACEs) remains high in patients with acute myocardial infarction (AMI) who undergo percutaneous coronary intervention (PCI), and early prediction models to guide their clinical management are lacking.

Objective: This study aimed to develop machine learning–based early prediction models for MACEs in patients with newly diagnosed AMI who underwent PCI.

Methods: A total of 1531 patients with AMI who underwent PCI from January 2018 to December 2019 were enrolled in this consecutive cohort. The data comprised demographic characteristics, clinical investigations, laboratory tests, and disease-related events. Four machine learning models—artificial neural network (ANN), k-nearest neighbors, support vector machine, and random forest—were developed and compared with the logistic regression model. Our primary outcome was the model performance that predicted the MACEs, which was determined by accuracy, area under the receiver operating characteristic curve, and F1-score.

Results: In total, 1362 patients were successfully followed up. With a median follow-up of 25.9 months, the incidence of MACEs was 18.5% (252/1362). The area under the receiver operating characteristic curve of the ANN, random forest, k-nearest neighbors, support vector machine, and logistic regression models were 80.49%, 72.67%, 79.80%, 77.20%, and 71.77%, respectively. The top 5 predictors in the ANN model were left ventricular ejection fraction, the number of implanted stents, age, diabetes, and the number of vessels with coronary artery disease.

Conclusions: The ANN model showed good MACE prediction after PCI for patients with AMI. The use of machine learning–based prediction models may improve patient management and outcomes in clinical practice.

(JMIR Form Res 2024;8:e48487) doi:10.2196/48487

KEYWORDS
acute myocardial infarction; percutaneous coronary intervention; machine learning; early prediction; cardiovascular event
Introduction

Acute myocardial infarction (AMI) is a common clinical acute and severe disease with rapid onset, rapid progression, and high mortality [1-3]. In 2017, there were approximately 695,000 new cases of AMI in the United States, and it is estimated that 325,000 people will have recurrent events [4]. There are approximately 500,000 new cases of AMI in China every year, and 2.5 million patients have a history of myocardial infarction [5]. As technology has advanced, percutaneous coronary intervention (PCI) has become the primary approach for treating AMI. Although PCI can significantly reduce the fatality rate of AMI, the rate of major adverse cardiovascular events (MACEs) among patients after PCI is still very high, which seriously affects the clinical outcomes of patients [6-10]. A study by Copeland-Halperin et al [11] showed that the incidence of MACEs in patients with AMI one year after PCI was 17.8% [11].

Identifying patients with AMI undergoing PCI who are at high risk of MACEs may help clinical decision-making incorporate timely measures to improve clinical outcomes. Some studies, such as Global Registry of Acute Coronary Event [12], Thrombolysis in Myocardial Infarction Risk [13,14], and Acute Catheterization and Urgent Intervention Triage Strategy-PCI [15], as well as studies that generated the Mayo Clinic PCI Risk and the China Acute Myocardial Infarction scoring systems, have explored the risks after PCI [16]. Despite these advances, individualized prediction of MACEs remains challenging with low specificity and positive predictive accuracy, and most of the methods rely on traditional parameter models, such as logistic regression, to screen for variables and build a series of risk-scoring models.

In recent years, machine learning methods that rely on a strong self-learning capability, such as random forest (RF), k-nearest neighbors (KNN), support vector machine (SVM), and artificial neural network (ANN) have become increasingly prevalent in prognostic prediction [1,13,17,18]. By calling various functions, these models can extract and integrate information from all kinds of complex data to make better predictions. A study of a consecutive cohort of patients with hypertrophic cardiomyopathy (HCM) presented a machine learning–based model to identify individual patients with HCM at high risk of developing advanced heart failure symptoms. The results showed that the 5-year risk prediction of progressive heart failure in patients with HCM can be estimated [19].

We found that machine learning models, such as RF, ANN, SVM, and KNN, perform well in clinical prognosis prediction research. Thus, this study sought to develop a machine learning–based model, integrating clinical, anatomical, and laboratory features, to predict MACEs in patients who have recently been diagnosed with AMI after their first PCI and improve overall patient outcomes by implementing earlier management.

Methods

Study Design, Setting and Participant Selection

This retrospective cohort study was conducted at the Department of Cardiovascular Medicine, the Second Affiliated Hospital of Nanchang University (a teaching tertiary hospital), in Jiangxi Province, China. We collected electronic medical records of patients with AMI who underwent PCI for the first time from January 2018 to December 2019. These patients were followed up through December 2021.

The inclusion criteria of the participants were as follows:

- The patient was ≥18 years of age.
- This was the patient’s first clinically diagnosed AMI (clinical evidence of AMI as evident from the detection of a rise or fall of cardiac troponin values and at least one of the following symptoms of myocardial ischemia: symptoms of acute myocardial ischemia, new ischemic electrocardiogram (ECG) changes, and development of pathological Q waves.
- PCI was performed for the first time at this hospital.
- Among the left main artery, left circumflex branch, left anterior descending branch, and right coronary artery, at least one had stenosis ≥50%.
- Complete medical records and follow-up data were available.

The following exclusion criteria were applied:

- History of PCI and coronary artery bypass grafting treatment
- Complications from other heart diseases requiring surgical procedures, such as heart bypass
- Recent active bleeding
- An intracerebral mass or an aneurysm

We adopted the “Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research” to guide the reporting of our study [20].

Data Collection, Definition of Outcomes, and Predictor Variables

Data were collected from electronic health records, including demographic characteristics, clinical investigations, the first laboratory tests, and disease-related events. MACEs were defined as cardiomyopathies (excluding infectious, familial, alcohol, and drug-related cardiomyopathies), hypertensive heart disease, recurrent myocardial infarction, heart failure, sudden cardiac death, revascularization, malignant arrhythmia, and stent thrombosis [21]. Abnormal Q waves were identified by the clinician based on ECG results. Left ventricular ejection fraction (LVEF) was defined as normal (more than 50%), mildly abnormal (40% to 50%), moderately abnormal (30% to 40%), and severely abnormal (less than 30%) [22]. According to the number of diseased coronary vessels and implanted stents, they were classified as I, II, III, and IV.
Ethics Approval
This study was reviewed and approved by the Second Affiliated Hospital of Nanchang University Medical Ethics Committee (No. Review 2017 No. (098)).

Data Preprocessing for Machine Learning Model Development
All analyses were performed with R software (version 4.0.1; R Core Team). The patients were randomly assigned to training (n=953, 70%) and testing (n=409, 30%) data sets by calling the createDataPartition function using the random number method, and chi-square tests showed that there was no statistical difference between them ($\chi^2=2.169; P=.14$). We developed machine learning models using the training data set. We analyzed the missing and out-of-range values with imputation methods. We used multiple imputation with chained equations to assign any missing predictor values [23]. The imputation processes were performed separately in the training and testing sets after the data were split. To improve the accuracy of the machine learning models and increase the speed of finding the optimal solution by gradient descent, we standardized and normalized all input variables before the model was built. To alleviate the problem of imbalanced classification samples, we adopted the random oversampling method. We used the ROSE package in R to generate new balanced training data. After random oversampling, the number of patients with MACE in the training data sets changed from 186 to 471.

Predictor Selection for Model Development
The model was built using demographic information (age and sex), personal comorbidities (diabetes and peripheral arterial disease), preoperative PCI (LVEF, the number of diseased vessels, and abnormal Q waves), serological examination (beta 2 microglobulin, B-type brain natriuretic peptide, glucose, serum creatinine clearance, and estimated glomerular filtration rate), and the characteristics of PCI (the number of implanted stents; n=65; Table S1 in Multimedia Appendix 1). A total of 12 variables with significant differences in the univariate analysis were included in the model development (Table 1).
Table 1. Baseline characteristics of the study patients (N=1362).

<table>
<thead>
<tr>
<th>Variables</th>
<th>MACEa (n=252)</th>
<th>Non-MACE (n=1110)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;65</td>
<td>101 (40.08)</td>
<td>543 (48.92)</td>
<td>.04</td>
</tr>
<tr>
<td>65</td>
<td>94 (37.30)</td>
<td>332 (29.91)</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>57 (22.62)</td>
<td>235 (21.17)</td>
<td></td>
</tr>
<tr>
<td>Diabetes, n (%)</td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Yes</td>
<td>75 (29.76)</td>
<td>261 (23.51)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>177 (70.24)</td>
<td>849 (76.49)</td>
<td></td>
</tr>
<tr>
<td>Vascular disease, n (%)</td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Yes</td>
<td>111 (44.05)</td>
<td>569 (51.26)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>141 (55.95)</td>
<td>541 (48.74)</td>
<td></td>
</tr>
<tr>
<td>Abnormal Q wave, n (%)</td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Yes</td>
<td>125 (49.60)</td>
<td>480 (43.24)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>127 (50.40)</td>
<td>630 (56.76)</td>
<td></td>
</tr>
<tr>
<td>LVEFb, n (%)</td>
<td></td>
<td></td>
<td>.005</td>
</tr>
<tr>
<td>&gt;50%</td>
<td>167 (66.27)</td>
<td>832 (74.95)</td>
<td></td>
</tr>
<tr>
<td>40%-50%</td>
<td>57 (22.62)</td>
<td>188 (16.94)</td>
<td></td>
</tr>
<tr>
<td>30%-40%</td>
<td>19 (7.54)</td>
<td>65 (5.86)</td>
<td></td>
</tr>
<tr>
<td>&lt;30%</td>
<td>9 (3.57)</td>
<td>25 (2.25)</td>
<td></td>
</tr>
<tr>
<td>Vessels with coronary artery disease, n (%)</td>
<td>45 (17.86)</td>
<td>288 (25.95)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>I</td>
<td>75 (29.76)</td>
<td>370 (33.33)</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>123 (48.81)</td>
<td>418 (37.66)</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>9 (3.57)</td>
<td>34 (3.06)</td>
<td></td>
</tr>
<tr>
<td>Implanted stent number, n (%)</td>
<td>10 (3.97)</td>
<td>40 (3.60)</td>
<td>.004</td>
</tr>
<tr>
<td>No stent</td>
<td>106 (42.06)</td>
<td>594 (53.51)</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>84 (33.33)</td>
<td>301 (27.12)</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>37 (14.68)</td>
<td>114 (10.27)</td>
<td></td>
</tr>
<tr>
<td>≥IV</td>
<td>15 (5.95)</td>
<td>61 (5.50)</td>
<td></td>
</tr>
<tr>
<td>Brain natriuretic peptide (pg/μL), mean (SD)</td>
<td>684.36 (997.90)</td>
<td>518.27 (773.65)</td>
<td>.01</td>
</tr>
<tr>
<td>Serum creatinine clearance (mL/min), mean (SD)</td>
<td>65.19 (30.18)</td>
<td>71.87 (44.35)</td>
<td>.02</td>
</tr>
<tr>
<td>EGFRc (ml/min), mean (SD)</td>
<td>75.68 (28.92)</td>
<td>80.55 (31.82)</td>
<td>.03</td>
</tr>
<tr>
<td>Beta 2 microglobulin (mg/L), mean (SD)</td>
<td>3.23 (3.61)</td>
<td>2.72 (5.51)</td>
<td>.03</td>
</tr>
<tr>
<td>Glucose (mmol/L), mean (SD)</td>
<td>7.22 (3.32)</td>
<td>6.68 (3.00)</td>
<td>.02</td>
</tr>
</tbody>
</table>

aMACE: major adverse cardiovascular events.
bLVEF: left ventricular ejection fraction.
cEGFR: estimated glomerular filtration rate.

Model Testing and Performance Evaluation

Based on a previous application of the model [24], the parameter range of the model was preset, and the GridSearchCV function was used to select the optimal parameters of each machine learning model.

To minimize potential overfitting in the above machine learning models, we called the trainControl function in the caret package of R language for 7-fold cross-validation during the development process. The model performance was assessed for accuracy, recall, precision, area under the receiver operating characteristic curve (AUC), and $F_1$-score in the testing data set. We identified
the important predictors through importance analysis of the variables. Logistic regression analysis was used to compare the absolute value of the coefficients of variables; RF was used to measure the importance of features by calculating information gain through entropy; and the ANN method was used to calculate the relative importance of variables based on the generalized weight method.

**Statistical Analysis**

The following R packages for machine learning approaches were used: caret, randomForest, and neuralnet. Baseline characteristics were compared with the Wilcoxon rank sum test for continuous variables and the chi-square test for categorical variables. We considered $P<.05$ (2-sided) to be statistically significant.

**Results**

A total of 1531 patients were screened; 140 patients who did not undergo PCI for the first time were excluded; 19 patients were lost to follow-up; and 1362 patients who were successfully followed up were included in this analysis (Figure 1). The mean follow-up time was 28.0 (SD 11.0) months (median 29.9 months). A total of 252 MACEs were observed, including 128 cases of recurrent myocardial ischemia and 117 cases of myocardial infarction and reinfarction. The positive rates of MACEs were 4.63%, 11.38%, 14.54%, and 18.50% at 30 days, 6 months, 1 year, and 3 years after PCI, respectively. MACEs occurred in 203 (18.7%) male patients and 49 (17.8%) female patients. As shown in Figure 2, the survival rate of the sample population decreased rapidly in the first 3 months after PCI, especially 30 days after PCI, and there was no difference in the log-rank test of the survival curve between male and female patients.

**Table 1** shows the baseline characteristics of the MACE group and the non-MACE group. Age, diabetes, peripheral and cerebrovascular history, LVEF, abnormal Q wave, the number of vessels with coronary artery disease, the number of implanted stents, brain natriuretic peptide, serum creatinine, estimated glomerular filtration rate, beta 2 microglobulin, and glucose were significantly different between the 2 groups ($P<.001$). The nonsignificant differences in variables between the 2 groups are shown in Table S1-S6 in Multimedia Appendix 1.

**Table 2** shows the performance of the 3 models with 7-fold cross-validation. ANN, KNN, SVM, RF, and logistic regression exhibited the best to worst performance in terms of their AUC, accuracy, recall, and $F_1$-score. However, KNN performed best in terms of precision. The average accuracy, recall, precision, AUC, and $F_1$-score of the ANN model were 80.52%, 81.33%, 69.94%, 83.68%, and 79.47%, respectively.

In the testing data set, the ANN model showed a higher AUC than RF and logistic regression. Figure 3 shows that the AUCs of the 3 models were 0.805, 0.798, 0.772, 0.727, and 0.718, respectively: the average accuracy for the above 3 models was 0.821, 0.741, and 0.729, respectively, and the average $F_1$-scores were 0.804, 0.722, and 0.709, respectively.

The 10 most important predictors in the ANN model are shown in **Table 3**. These were LVEF (0.27), the number of implanted stents (0.14), age (0.13), diabetes (0.10), the number of vessels with coronary artery disease (0.09), vascular disease (0.08), brain natriuretic peptide (0.05), glucose (0.05), beta 2 microglobulin (0.04), and abnormal Q wave (0.02).

![Figure 1. Flowchart for patient enrollment. AMI: acute myocardial infarction; MACE: major adverse cardiovascular event.](https://formative.jmir.org/2024/1/e48487)
Figure 2. Prognostic survival curve of patients with acute myocardial infarction undergoing percutaneous coronary intervention.

Table 2. Comparison of models for predicting major adverse cardiovascular events based on 7-fold cross-validation.

<table>
<thead>
<tr>
<th>Models</th>
<th>Accuracy, mean (SD)</th>
<th>Recall, mean (SD)</th>
<th>Precision, mean (SD)</th>
<th>AUC&lt;sup&gt;a&lt;/sup&gt;, mean (SD)</th>
<th>F&lt;sub&gt;1&lt;/sub&gt;-score, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression</td>
<td>72.37 (2.05)</td>
<td>67.33 (8.42)</td>
<td>59.62 (8.34)</td>
<td>73.52 (2.37)</td>
<td>71.11 (6.01)</td>
</tr>
<tr>
<td>K-nearest neighbors</td>
<td>81.44 (2.22)</td>
<td>80.23 (1.56)</td>
<td>70.22 (7.23)</td>
<td>81.87 (3.32)</td>
<td>77.95 (5.70)</td>
</tr>
<tr>
<td>Support vector machine</td>
<td>74.91 (3.03)</td>
<td>80.03 (1.76)</td>
<td>65.94 (7.02)</td>
<td>78.68 (1.82)</td>
<td>76.41 (5.92)</td>
</tr>
<tr>
<td>Random forest</td>
<td>73.44 (1.58)</td>
<td>71.23 (1.56)</td>
<td>61.22 (7.23)</td>
<td>74.87 (2.12)</td>
<td>71.92 (6.30)</td>
</tr>
<tr>
<td>Artificial neural network</td>
<td>80.52 (1.13)</td>
<td>81.33 (0.56)</td>
<td>69.94 (7.02)</td>
<td>83.68 (1.82)</td>
<td>79.47 (4.57)</td>
</tr>
</tbody>
</table>

<sup>a</sup>AUC: area under the receiver operating characteristic curve.

Figure 3. The area under the receiver operating characteristic (ROC) curve of artificial neural network (ANN), random forest (RF), k-nearest neighbors (KNN), support vector machine (SVM), and logistic regression models.
Discussion

Principal Findings

In this study, we developed a machine learning–based model integrating clinical, anatomical, and laboratory test features to predict MACEs in patients with newly diagnosed AMI after their first PCI. The major findings suggest that the ANN model had higher predictive accuracy (accuracy of 87.99%, AUC of 0.81, and F1-score of 0.71), compared to RF, KNN, SVM, and logistic regression.

Among the patients with AMI in this study, the rates of MACEs at 30 days, 6 months, 1 year, and 3 years after PCI were 4.63%, 11.38%, 14.54%, and 18.50%, respectively, and the incidence of MACEs at 30 days after PCI was slightly less than the 5.5% reported in the Harmonizing Outcomes with RevascularIZatiON and Stents in Acute Myocardial Infarction study (HORIZONS-AMI) [25]. The incidence of MACEs at half a year was higher than the 6.67% reported by Chow et al [26], consistent with the 2-year rate of MACEs reported by Sanmenxia City (18.06%). The survival condition of patients with AMI after PCI was slightly different from that in other studies. The participants in this study were all patients who were first diagnosed with AMI and underwent PCI for the first time, and their prognosis was better than that of patients with previous myocardial infarctions and multiple PCIs [27]. In addition, the progression of a patient’s disease is affected by not only individual differences but also access to medical resources and services. The HORIZONS-AMI trial was first reported in 2012. Although the treatment level in the HORIZONS-AMI trial was higher than that available in China at that time, with the development of China’s economy, the progress of science and technology, and the substantial improvement of medical care, the MACE rate obtained in our study was lower than that reported in the HORIZONS-AMI study.

One study found that machine learning demonstrated the highest performance for risk prediction in patients with extracardiac vascular disease for the prediction of both arrhythmogenic cardiomyopathy and MACEs [10]. McCord et al [28] proposed that machine learning can be used to assess AMI within 30 minutes and that the algorithm has high diagnostic and prognostic utility. In this study, 3 algorithms were used to predict MACE occurrence for patients with newly diagnosed AMI undergoing PCI treatment for the first time. The MACE prediction ability of the logistic regression model was lower than that of the ANN model and almost the same as that of the RF model. However, the positive predictive values of these 3 prediction models were not high. Kuang et al [29] also found that the ANN model had the best predictive value for the transition from mild cognitive impairment to Alzheimer disease with ideal stability [29]. The positive predictive values of the RF model and the logistic regression model were both approximately 50%, which means that their predictive ability for MACEs was poor. Their shortcomings may be associated with class imbalances [30], which can easily cause the predicted results to be biased toward a large number of classes (the positive type of fault can be placed into the negative class). ANNs, with their powerful self-adaptability, self-organization, fault tolerance, and “black box” operation of nonlinear mapping, are especially suitable for solving problems with complex internal mechanisms and have been widely used in various disciplines [31].

Our results indicated that the 3-year prognostic risk among patients with AMI undergoing their first PCI was mainly related to age, ECG characteristics, ventricular ejection ability, coronary artery lesions, stent implantation after PCI, and some serological variables. Yang et al [32] found that the risk ratio of hospital deaths after PCI was 3.723 (95% CI 2.86-4.84) for South Korean patients aged >65 years relative to those aged ≤65 years. A Korean multicenter AMI National Institutes of Health–registered project found that the MACE rate, 3 years after PCI, among patients with AMI with an LVEF <40% was 3.34 times that of the control group [33]. Fam et al [34] conducted a retrospective study on patients with clinical AMI in Asian multiethnic groups and found that the risk of MACEs among patients with diabetes, 2 years after PCI, was 1.84 times higher than that among patients without diabetes [34]. Diabetes is a chronic metabolic disease, and long-term diabetes is often accompanied by bleeding disorders, vascular endothelial dysfunction, small artery lesions, high blood sugar [35], hemostatic disorders [36], endothelial dysfunction, and a series of other changes [37]. These characteristics will accelerate the process of atherosclerotic disease deterioration. The number of coronary artery lesions

### Table 3. Importance of each variable in the artificial neural network model.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left ventricular ejection fraction</td>
<td>0.27</td>
</tr>
<tr>
<td>The number of implanted stents</td>
<td>0.14</td>
</tr>
<tr>
<td>Age</td>
<td>0.13</td>
</tr>
<tr>
<td>Diabetes</td>
<td>0.10</td>
</tr>
<tr>
<td>The number of vessels with coronary artery disease</td>
<td>0.09</td>
</tr>
<tr>
<td>Vascular disease</td>
<td>0.08</td>
</tr>
<tr>
<td>Brain natriuretic peptide</td>
<td>0.05</td>
</tr>
<tr>
<td>Glucose</td>
<td>0.05</td>
</tr>
<tr>
<td>Beta 2 microglobulin</td>
<td>0.04</td>
</tr>
<tr>
<td>Abnormal Q wave</td>
<td>0.02</td>
</tr>
</tbody>
</table>

https://formative.jmir.org/2024/1/e48487
and the number of stents implanted in a patient are also positively correlated with the risk of postoperative MACEs to a certain extent. This may be because a higher number of vessels with coronary artery disease and the number of implanted stents tend to indicate a more serious condition, leading to a worse prognosis for the patients. Hongbo et al [38] found that the probability of a poor prognosis in patients with multiple coronary artery lesions was 20.0%, compared with 6.98% in patients with single coronary artery lesions [38].

The results of the machine learning model showed that predictors like LVEF, number of implanted stents, and age were more important to the model. LVEF is a common variable that reflects left ventricular function, and patients with a low LVEF have a significantly higher MACE rate [39]. An increase in age can lead to the aggravation of atherosclerosis [40]. The number of implanted stents may be related to the severity of the disease and the extent of the infarction [41]. This reminds us that we should pay special attention to the prognosis of patients with AMI who have a low LVEF value, older age, and more implanted stents in clinical practice.

Study Limitations
This study has some limitations. First, there may have been an issue of survival bias in the study, as patients with missing follow-up data were excluded. Second, the data have missing values. We have filled missing values with multiple imputation; however, imputation with these techniques could synthetically reduce the variance in these variables and may have affected the accuracy of the constructed model. Finally, although the models were internally validated with data from the same hospital, further work should include validation with external data from other hospitals or centers.

Conclusions
This study revealed that the ANN model showed good MACE prediction performance for patients with AMI after PCI, and it identified the most important predictors, which may aid in clinical decision-making and improve outcomes. This model needs to be externally validated in larger populations and multicenter settings.

Acknowledgments
The authors are grateful to the following individuals for contributing data: Li Li, Nan Wang, LiFang Deng, ShanLan Yang, and Bin Liu. This study was supported by the National Natural Science Foundation of China (grants 81960611, 82160645, 82360667, and 81260444), Natural Science Foundation of Jiangxi Province (grants 20192BAA208005 and 20212BAB206091), and the National Undergraduate Training Program for Innovation and Entrepreneurship (grant 202010403019).

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Additional statistics.

References


Abbreviations

AMl: acute myocardial infarction
ANN: artificial neural network
AUC: area under the receiver operating characteristic curve
ECG: electrocardiogram
HCM: hypertrophic cardiomyopathy
HORIZONS-AMI: Harmonizing Outcomes with RevascularIZatiON and Stents in Acute Myocardial Infarction
KNN: k-nearest neighbors
LVEF: left ventricular ejection fraction
MACE: major adverse cardiovascular event
PCI: percutaneous coronary intervention
RF: random forest
SVM: support vector machine

https://formative.jmir.org/2024/1/e48487
Differences in Psychological Inflexibility Among Men With Erectile Dysfunction Younger and Older Than 40 Years: Web-Based Cross-Sectional Study

Junichi Saito1, PhD; Hiroaki Kumano2, MD, PhD; Mohammad Ghazizadeh3, MD, PhD; Chigusa Shimokawa3, MA; Hideki Tanemura3, MBA

1Comprehensive Research Organization, Waseda University, Saitama, Japan
2Faculty of Human Sciences, Waseda University, Saitama, Japan
3Logos Science Corp, Tokyo, Japan

Corresponding Author:
Junichi Saito, PhD
Comprehensive Research Organization
Waseda University
2-579-15, Mikajima, Tokorozawa
Saitama, 359-1192
Japan
Phone: 81 429498113
Email: tekuteke@aoni.waseda.jp

Abstract

Background: Psychological inflexibility is a core concept of acceptance and commitment therapy (ACT), which is a comprehensive, transdiagnostic interpretation of mental health symptoms. Erectile dysfunction (ED) is a condition that affects male sexual performance, involving the inability to achieve and maintain a penile erection sufficient for satisfactory sexual activity. Psychosocial factors primarily influence ED in men younger than 40 years, whereas biological factors are more likely to be the underlying cause in older men.

Objective: This web-based cross-sectional study examined differences in depression, anxiety, and psychological inflexibility among men with ED younger and older than 40 years in a Japanese population.

Methods: We used a web-based survey to gather data from various community samples. ED was assessed by the International Index of Erectile Function - 5 (IIEF-5) questionnaire, while depression, anxiety, and psychological inflexibility were evaluated by the Patient Health Questionnaire-9 (PHQ-9), General Anxiety Disorder-7 (GAD-7), Acceptance and Action Questionnaire-II (AAQ-II), Cognitive Fusion Questionnaire (CFQ), and Valuing Questionnaire–Obstacle Subscale (VQ-OB) questionnaires. The chi-square test estimated the scores of PHQ-9 and GAD-7 among men with ED, comparing those younger than 40 years and those older than 40 years. Additionally, a two-way ANOVA was conducted with ED severity and age group as independent variables, assessing psychological inflexibility.

Results: Valid responses from 643 individuals (mean age 36.19, SD 7.54 years) were obtained. Of these, 422 were younger than 40 years (mean age 31.76, SD 5.00 years), and 221 were older than 40 years (mean age 44.67, SD 2.88 years). There was a statistical difference in the prevalence of depression as judged by PHQ≥10 between men with ED younger and older than 40 years (P<.001). On the other hand, there was no difference in the prevalence of anxiety as judged by GAD≥10 (P=.12). The two-way ANOVA revealed that the interactions for CFQ (P=.04) and VQ-OB (P=.01) were significant. The simple main effect was that men with ED younger than 40 years had significantly higher CFQ (P=.01; d=0.62) and VQ-OB (P<.001; d=0.87) scores compared to those older than 40 years in moderate ED and severe ED. Additionally, it was found that men younger than 40 years with moderate to severe ED had significantly higher CFQ (P=.01; d=0.42) and VQ-OB (P=.02; d=0.38) scores compared to men younger than 40 years without ED. On the other hand, no interaction was found for AAQ-II (P=.16) scores.

Conclusions: To the best of our knowledge, this web-based cross-sectional study is the first to examine the relationship between psychological inflexibility and ED. We conclude that men with moderate and severe ED younger than 40 years have higher psychological inflexibility and might be eligible for ACT.

(JMIR Form Res 2024;8:e45998) doi:10.2196/45998
KEYWORDS
erectile dysfunction; acceptance and commitment therapy; psychological inflexibility; depression; anxiety; men; cross-sectional study; psychological; utility; psychosocial; men; therapy; impotence; erection

Introduction
The efficacy of acceptance and commitment therapy (ACT) has been evaluated in numerous randomized controlled studies exploring various targeted conditions [1]. There is supporting evidence for ACT across various physical illnesses (eg, chronic pain [2], diabetes [3], epilepsy [4], cancer [5], and irritable bowel syndrome [6]). Many of these studies use a transdiagnostic method to analyze psychological issues within individual health conditions. Psychological inflexibility is a core concept of ACT, which is a comprehensive, transdiagnostic interpretation of mental health symptoms [7]. Psychological inflexibility highlights two interrelated processes: cognitive fusion and experiential avoidance. Cognitive fusion represents the phenomenon by which individuals are influenced by the literal meaning of their thoughts instead of viewing them as transient internal states [8]. Experiential avoidance represents an attempt or desire to suppress unwanted internal experiences, such as emotions, thoughts, memories, and bodily sensations [9]. These processes are obstacles to one’s valued living activities, decreasing well-being [10].

Erectile dysfunction (ED) is a condition that affects male sexual performance, involving the inability to achieve and maintain a penile erection sufficient for satisfactory sexual activity [11]. Several reviews and clinical guidelines are available for ED. However, many of these approaches to assessing and treating ED purely from a medical perspective seldom address the psychosocial components of ED [12]. Pharmacological treatment alone does not respond to all the concurrent factors of ED, including anxiety, loss of self-confidence, depressed mood, difficulties in a couple’s communication, relationship disputes, or a partner’s sexual problems [13]. Recent systematic reviews have shown that combining phosphodiesterase-5 inhibitors with psychological treatment exhibits significant potential for treating ED [14].

There is a widespread assumption that psychosocial factors primarily influence ED in men younger than 40 years, whereas biological factors are more likely to be the underlying cause of ED in older men. Moore et al [15] showed different symptom patterns among patients with ED according to age groups. They reported that younger men had comparatively more significant depressive symptoms, with lower relationship satisfaction, more negative reactions from partners, and lower job satisfaction. Given these findings, it is possible that people younger than 40 years are more psychologically inflexible than those older than 40 years and that ACT is more effective for them. However, studies on ACT for ED remain limited, with only a few identified. Therefore, this cross-sectional study used assessments to evaluate depression, anxiety, and psychological inflexibility in men younger and older than 40 years. However, it is well known that ED prevalence varies across geographical groups [16,17]; therefore, it is essential to research ED etiology according to different racial, cultural, religious, and socioeconomic backgrounds. There might be many potential patients with ED in Japan, so we conducted a web-based survey for this study.

Methods
Participants
To gather data from a wide range of community samples, we used a web-based survey, conducted with the assistance of a marketing research service provider (Rakuten Insight, Inc) in Japan. Based on the International Index of Erectile Function – 5 (IIEF-5) cutoff point [18], participants of all severities were recruited to include a certain percentage of patients of all ages. All enrolled participants followed the following criteria: (1) male; (2) aged 20 to 50 years; and (3) married or living with a fixed sexual partner for more than 6 months. The exclusion criteria were as follows: (1) sexual dysfunction caused by Peyronie disease or other organic lesions of the external genitalia; (2) prostate cancer, hypertensive disease, cardiac disease, cerebrovascular disease, chronic kidney disease, and diabetes; and (3) a history of sertraline or other medicines that may influence erection and psychological symptoms.

Participants were first instructed that this survey would be administered anonymously, and their responses were not compulsory. Then, those participants who agreed to participate in this research responded to the surveys. Participants were given points to exchange for items within the survey company’s system as a reward.

Ethical Considerations
This study was approved by the Waseda University Academic Research Ethical Review Committee (2019-363). The study protocol followed the guidelines for epidemiological studies in accordance with the Declaration of Helsinki.

Measurements

International Index of Erectile Function-5 (IIEF-5)
The Japanese version of IIEF-5 is a 5-item self-report questionnaire designed to measure erectile function [18]. Items are rated on a 5-point Likert-type scale, ranging from 1 to 5. The total score can range from 5 to 25, with high scores meaning high erectile function. Based on the original validation studies, the total score can then be interpreted as suggesting “no ED” (22-25), “mild ED” (17-21), “mild-to-moderate ED” (12-16), “moderate ED” (8-11), and “severe ED” (5-7).

Patient Health Questionnaire-9 (PHQ-9)
The Japanese version of the Patient Health Questionnaire-9 (PHQ-9) is a 9-item self-report questionnaire designed to measure depression [19]. Items are rated on a 4-point Likert-type scale, ranging from 0 to 3. The total score can range from 0 to 27, with high scores meaning high depression. Based on the original validation studies, the total score can then be interpreted as suggesting no depression (0-4), mild depression (5-9), moderate depression (10-14), moderately severe depression (15-19), severe depression (20-27).

Measurement of Psychological Inflexibility
The Japanese version of the Psychological Inflexibility Scale [20] is a 7-item self-report questionnaire designed to measure psychological inflexibility. Items are rated on a 7-point Likert-type scale, ranging from 1 (completely disagree) to 7 (completely agree). The total score can range from 7 to 49, with high scores meaning high psychological inflexibility.
(15-19), or severe depression (20-27). A cutoff score of 10 is suggested as indicating a possible diagnosis of depressive disorder.

**Generalized Anxiety Disorder-7 (GAD-7)**

The Japanese version of the Generalized Anxiety Disorder-7 (GAD-7) questionnaire is a 7-item self-report questionnaire designed to measure generalized anxiety disorder [20]. Items are rated on a 4-point Likert-type scale, ranging from 0 to 3. The total score can range from 0 to 21, with high scores meaning high anxiety. Based on the original validation studies, the total score can then be interpreted as suggesting no anxiety (0-4), mild anxiety (5-9), moderate anxiety (10-14), or severe anxiety (14-21). A cutoff score of 10 is suggested as indicating a possible diagnosis of generalized anxiety disorder.

**Acceptance and Action Questionnaire-II (AAQ-II)**

The Japanese version of the Acceptance and Action Questionnaire-II (AAQ-II) is a 7-item self-report questionnaire designed to measure experiential avoidance [21]. Items are rated on a 7-point Likert-type scale, ranging from 1 to 7. The total score can range from 7 to 49, with high scores meaning high experiential avoidance.

**Cognitive Fusion Questionnaire (CFQ)**

The Japanese version of the Cognitive Fusion Questionnaire (CFQ) is a 7-item self-report questionnaire designed to measure cognitive fusion [22]. Items are rated on a 7-point Likert-type scale, ranging from 1 to 7. The total score can range from 7 to 49, with high scores meaning high cognitive fusion.

**Valuing Questionnaire–Obstacle Subscale (VQ-OB)**

The Japanese version of the Valuing Questionnaire–Obstacle Subscale (VQ-OB) is a 5-item self-report questionnaire designed to measure obstruction of valued living [23]. Items are rated on a 7-point Likert-type scale, ranging from 0 to 6. The total score can range from 0 to 30, with high scores meaning high obstruction of valued living.

**Statistical Analysis**

We used mean (SD) values to describe numerical data and counts and percentages to describe categorical data. The chi-square tests estimated categorical data, and numerical data were estimated by t tests. A two-way ANOVA test was used to assess the differences in men with ED aged younger and older than 40 years regarding psychological inflexibility and the interaction between them. Post hoc tests were conducted using the Holm method to control for type I errors. Cohen d index was calculated as effect sizes, serving as standardized indicators unaffected by sample sizes. All tests were 2-tailed, and a statistical difference was assumed when the P value was <.05. All statistical analyses were conducted through IBM SPSS Statistics (version 25.0; IBM Corp).

**Results**

We obtained valid responses from 643 individuals (mean age 36.19, SD 7.54 years). Of these, 422 were younger than 40 years (mean age 31.76, SD 5.00 years), and 221 were older than 40 years (mean age 44.67, SD 2.88 years). Table 1 shows the demographic characteristics of participants by age difference. No statistical difference was found in ED severity, phosphodiesterase-5 inhibitors use, and marriage status between men with ED younger and older than 40 years.

### Table 1. Demographic characteristics of men with erectile dysfunction (ED) younger and older than 40 years.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>&lt;40 years of age (n=422)</th>
<th>≥40 years of age (n=221)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>31.76 (5.00)</td>
<td>44.67 (2.88)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
<td></td>
<td>.44</td>
</tr>
<tr>
<td>Single</td>
<td>77 (18.25)</td>
<td>35 (15.84)</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>345 (81.75)</td>
<td>186 (84.16)</td>
<td></td>
</tr>
<tr>
<td>Duration of marriage (years), mean (SD)</td>
<td>4.87 (3.90)</td>
<td>10.43 (7.53)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>IIEF-5a severity, n (%)</td>
<td></td>
<td></td>
<td>.84</td>
</tr>
<tr>
<td>No ED</td>
<td>84 (19.91)</td>
<td>49 (22.17)</td>
<td></td>
</tr>
<tr>
<td>Mild to mild-to-moderate ED</td>
<td>226 (53.55)</td>
<td>112 (50.68)</td>
<td></td>
</tr>
<tr>
<td>Moderate to severe ED</td>
<td>112 (26.54)</td>
<td>60 (27.15)</td>
<td></td>
</tr>
<tr>
<td>PDE-5b use, n (%)</td>
<td></td>
<td></td>
<td>.63</td>
</tr>
<tr>
<td>Not using</td>
<td>337 (79.86)</td>
<td>180 (81.45)</td>
<td></td>
</tr>
<tr>
<td>Using</td>
<td>85 (20.14)</td>
<td>41 (18.55)</td>
<td></td>
</tr>
</tbody>
</table>

aIIEF-5: International Index of Erectile Function - 5.
bPDE-5: phosphodiesterase-5 inhibitor.

The prevalence of depression as judged by PHQ≥10 among men younger than 40 years was 39.81% (168/422), and it was 24.89% (55/221) among those older than 40 years. There was a statistical difference in the prevalence of depression between the two groups (P<.001). In addition, the prevalence of anxiety, as judged by GAD≥10, was 27.25% (115/422) among men...
younger than 40 years, and it was 21.72% (48/221) among those older than 40 years. There was no difference in the prevalence of anxiety between men with ED in the two age groups ($P=.12$). Table 2 illustrates these results.

Table 2. Prevalence of depression and anxiety among men with erectile dysfunction younger and older than 40 years.

<table>
<thead>
<tr>
<th>Questionnaires and characteristics</th>
<th>&lt;40 years of age (n=422)</th>
<th>≥40 years of age (n=221)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHQ-9a</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No depression or mild depression (PHQ-9&lt;10), n (%)</td>
<td>254 (60.19)</td>
<td>166 (75.11)</td>
<td></td>
</tr>
<tr>
<td>Prevalence of depression (PHQ-9≥10), n (%)</td>
<td>168 (39.81)</td>
<td>55 (24.89)</td>
<td></td>
</tr>
<tr>
<td>GAD-7b</td>
<td></td>
<td></td>
<td>.13</td>
</tr>
<tr>
<td>No anxiety or mild anxiety (GAD-7&lt;10), n (%)</td>
<td>307 (72.75)</td>
<td>173 (78.52)</td>
<td></td>
</tr>
<tr>
<td>Prevalence of anxiety (GAD-7≥10), n (%)</td>
<td>115 (27.25)</td>
<td>48 (21.72)</td>
<td></td>
</tr>
</tbody>
</table>

$^a$PHQ-9: Patient Health Questionnaire-9.
$^b$GAD-7: Generalized Anxiety Disorder-7.

The two-way ANOVA was performed with ED severity and age (<40 or ≥40 years) as independent variables and the scores of AAQ, CFQ, and VQ-OB as dependent variables. The results showed no significant differences in AAQ-II ($P=.14$), CFQ ($P=.08$), and VQ-OB ($P=.30$) scores attributed to ED severity. Moreover, no difference in ED severity or psychological inflexibility depending on the duration of the marriage was found. On the other hand, there were significant differences in the scores of CFQ ($P=.04$) and VQ-OB ($P=.004$) attributed to age. As the interactions were significant for CFQ ($P=.04$) and VQ-OB ($P=.01$) scores, the simple main effect was examined. It was found that men with ED younger than 40 years had significantly higher CFQ ($P=.01; d=0.62$) and VQ-OB ($P<.001; d=0.87$) scores compared to those older than 40 years, in cases of moderate and severe ED. Additionally, it was found that men with moderate to severe ED younger than 40 years had significantly higher CFQ ($P=.01; d=0.42$) and VQ-OB ($P=.02; d=0.38$) scores compared to men with no ED younger than 40 years. These results are illustrated in Table 3 and Figures 1 and 2.

Table 3. Two-way ANOVA results of the influence of erectile dysfunction (ED) severity, age, and interaction on psychological inflexibility.

<table>
<thead>
<tr>
<th>Parameters and factors</th>
<th>Sum of squares</th>
<th>Mean squares</th>
<th>$F$ test (df)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAQ-II$^a$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED severity</td>
<td>281.60</td>
<td>140.80</td>
<td>1.91 (2,637)</td>
<td>.14</td>
</tr>
<tr>
<td>Age</td>
<td>204.74</td>
<td>204.74</td>
<td>2.78 (1,637)</td>
<td>.10</td>
</tr>
<tr>
<td>ED severity x age</td>
<td>267.27</td>
<td>133.64</td>
<td>1.81 (2,637)</td>
<td>.16</td>
</tr>
<tr>
<td>CFQ$^b$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED severity</td>
<td>433.63</td>
<td>216.82</td>
<td>2.49 (2,637)</td>
<td>.08</td>
</tr>
<tr>
<td>Age</td>
<td>348.12</td>
<td>348.12</td>
<td>3.99 (1,637)</td>
<td>.04</td>
</tr>
<tr>
<td>ED severity x age</td>
<td>534.92</td>
<td>267.46</td>
<td>3.07 (2,637)</td>
<td>.04</td>
</tr>
<tr>
<td>VQ-OB$^c$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED severity</td>
<td>74.56</td>
<td>37.28</td>
<td>1.19 (2,637)</td>
<td>.30</td>
</tr>
<tr>
<td>Age</td>
<td>263.36</td>
<td>263.36</td>
<td>8.38 (1,637)</td>
<td>.004</td>
</tr>
<tr>
<td>ED severity x age</td>
<td>250.82</td>
<td>125.41</td>
<td>3.99 (2,637)</td>
<td>.01</td>
</tr>
</tbody>
</table>

$^a$AAQ-II: Acceptance and Action Questionnaire-II.
$^b$CFQ: Cognitive Fusion Questionnaire.
$^c$VQ-OB: Valuing Questionnaire–Obstacle Subscale.
**Discussion**

**Principal Findings**

This cross-sectional study evaluated depression, anxiety, and psychological inflexibility in men younger and older than 40 years with ED. There was no statistical difference in demographic characteristics between the two groups. The average age of the participants was 31.76 (SD 5.00) years in men younger than 40 years; the population was assumed to have mainly mild-to-moderate psychogenic ED. On the other hand, the average age of the participants was 44.67 (SD 2.88) years in men older than 40 years; the population was assumed to have mostly mild-to-moderate organic ED.

Depression was found in both groups. The results of our study were consistent with a previous study [15], which found that younger men had comparatively greater depressive symptoms. In contrast, the prevalence of anxiety was not different between the two age groups. One possible reason is that the anxiety in men with ED is not general anxiety but specific anxiety about sexual situations. Masters and Johnson [24] highlighted the...
Men with ED younger than 40 years had significantly higher CFQ and VQ-OB scores than those older than 40 years in cases of moderate and severe ED. Furthermore, men with moderate-to-severe ED younger than 40 years had significantly higher CFQ and VQ-OB scores compared to men without ED. These results partly support our hypothesis that men younger than 40 years are more psychologically inflexible than those older than 40 years. Cognitive fusion might be the critical component of ACT for ED. For example, the fusion with sexual performance anxiety, such as “I might fail again,” makes it impossible to pay attention to the sexual partner, which results in erectile failure. It is also consistent with Barlow’s theory [27]. Barlow [27] proposed a model for the interaction of anxiety and cognitive interference. This model examines how anxiety and cognitive interference interact, particularly in a sexual context, where a lack of control over one’s arousal diverts attention from erotic arousal to physical arousal and the negative consequences associated with failure to attain an erection.

On the other hand, there were no significant differences in the scores of the AAQ-II, which might be related to psychometric issues with AAQ-II. To date, the most used self-report measure of psychological inflexibility, especially experience avoidance, has been the AAQ-II. There was no significant difference in ED severity and psychological inflexibility depending on the duration of the marriage. However, various issues regarding the AAQ-II have emerged from the existing literature [28]. The authors found that the AAQ-II faced challenges in distinguishing psychological inflexibility from other related constructs. One of the main challenges is that the AAQ-II may not be specific enough to capture the unique aspects of psychological inflexibility in the context of ED. Another challenge is that the AAQ-II may not be sensitive enough to detect small changes in psychological inflexibility over time.

Conclusions
To the best of our knowledge, this web-based cross-sectional study was the first to examine the relationship between psychological inflexibility and ED. We conclude that men with moderate and severe ED younger than 40 years have higher psychological inflexibility and might be eligible for the ACT. In addition, developing a Japanese version of the questionnaire is necessary to measure ED-related anxiety and psychological inflexibility.

Data Availability
The data sets generated and analyzed during this study are not publicly available due to the Waseda University Academic Research Ethical Review Committee’s data-sharing policy but are available from the corresponding author upon reasonable request.

Conflicts of Interest
This research was funded by Logos Science Corp, Ltd, Tokyo, Japan. MG, CS, and HT are members of the Logos Science Corp, Ltd.

References


Abbreviations

AAQ-II: Acceptance and Action Questionnaire - II
ACT: Acceptance and Commitment Therapy
CFQ: Cognitive Fusion Questionnaire
ED: Erectile Dysfunction
GAD-7: Generalized Anxiety Disorder - 7
IIEF-5: International Index of Erectile Function - 5
PHQ-9: Patient Health Questionnaire - 9
VQ-OB: Valuing Questionnaire–Obstacle Subscale

©Junichi Saito, Hiroaki Kumano, Mohammad Ghazizadeh, Chigusa Shimokawa, Hideki Tanemura. Originally published in JMIR Formative Research (https://formative.jmir.org), 03.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Determining Distinct Suicide Attempts From Recurrent Electronic Health Record Codes: Classification Study

Kate H Bentley\textsuperscript{1,2*}, PhD; Emily M Madsen\textsuperscript{1,3*}, BSc; Eugene Song\textsuperscript{1,3}, BA; Yu Zhou\textsuperscript{1,3}, MPP; Victor Castro\textsuperscript{4}, MS; Hyunjoon Lee\textsuperscript{1,3}, MSc; Younga H Lee\textsuperscript{2,3}, PhD; Jordan W Smoller\textsuperscript{1,2,3}, MD, SCD

\textsuperscript{1}Center for Precision Psychiatry, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States
\textsuperscript{2}Department of Psychiatry, Harvard Medical School, Boston, MA, United States
\textsuperscript{3}Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, United States
\textsuperscript{4}Mass General Brigham Research Information Science and Computing, Somerville, MA, United States

* these authors contributed equally

Corresponding Author:
Kate H Bentley, PhD
Center for Precision Psychiatry
Department of Psychiatry
Massachusetts General Hospital
185 Cambridge Street
2nd Floor
Boston, MA, 02114
United States
Phone: 1 6177247741
Email: kbentley@mgh.harvard.edu

Abstract

Background: Prior suicide attempts are a relatively strong risk factor for future suicide attempts. There is growing interest in using longitudinal electronic health record (EHR) data to derive statistical risk prediction models for future suicide attempts and other suicidal behavior outcomes. However, model performance may be inflated by a largely unrecognized form of “data leakage” during model training: diagnostic codes for suicide attempt outcomes may refer to prior attempts that are also included in the model as predictors.

Objective: We aimed to develop an automated rule for determining when documented suicide attempt diagnostic codes identify distinct suicide attempt events.

Methods: From a large health care system’s EHR, we randomly sampled suicide attempt codes for 300 patients with at least one pair of suicide attempt codes documented at least one but no more than 90 days apart. Supervised chart reviewers assigned the clinical settings (ie, emergency department [ED] versus non-ED), methods of suicide attempt, and intercode interval (number of days). The probability (or positive predictive value) that the second suicide attempt code in a given pair of codes referred to a distinct suicide attempt event from its preceding suicide attempt code was calculated by clinical setting, method, and intercode interval.

Results: Of 1015 code pairs reviewed, 835 (82.3%) were nonindependent (ie, the 2 codes referred to the same suicide attempt event). When the second code in a pair was documented in a clinical setting other than the ED, it represented a distinct suicide attempt 3.3% of the time. The more time elapsed between codes, the more likely the second code in a pair referred to a distinct suicide attempt event from its preceding code. Code pairs in which the second suicide attempt code was assigned in an ED at least 5 days after its preceding suicide attempt code had a positive predictive value of 0.90.

Conclusions: EHR-based suicide risk prediction models that include International Classification of Diseases codes for prior suicide attempts as a predictor may be highly susceptible to bias due to data leakage in model training. We derived a simple rule to distinguish codes that reflect new, independent suicide attempts: suicide attempt codes documented in an ED setting at least 5 days after a preceding suicide attempt code can be confidently treated as new events in EHR-based suicide risk prediction models. This rule has the potential to minimize upward bias in model performance when prior suicide attempts are included as predictors in EHR-based suicide risk prediction models.
KEYWORDS
suicide; suicide attempt; self-injury; electronic health record; informatics; automated rule; psychiatry; machine learning

Introduction
Suicide is the tenth leading cause of death in the United States, with more than 48,000 suicide deaths annually [1]. Over the past 20 years, the suicide rate has increased by over 35% [2]. Most people who die by suicide have recently interacted with the health care system, with over half having a health care visit in the month prior to death [3,4]. Health care systems thus offer a key opportunity to identify people at high risk for suicide. Unfortunately, clinicians are poor at predicting who will make a suicide attempt [5] and traditionally studied risk factors perform no better than chance at predicting future suicidal behavior [6].

Recent work has focused on developing and validating machine learning models that use routinely collected electronic health record (EHR) data to predict future suicidal behavior [7]. Such models have demonstrated high levels of accuracy, exceeding that seen with clinician prediction and usual clinical risk factors [6,8-10]. EHR-based suicide risk prediction models, however, face one significant challenge that to date has not been adequately addressed. Suicide attempt is generally the outcome of interest in these models and is typically defined by International Classification of Diseases (ICD) diagnostic codes [11,12]. Within a given patient’s EHR, a suicide attempt code may be given multiple times across distinct health care encounters, often over very short periods of time (eg, days and weeks). Such “recurrent” codes may represent either distinct, new events (ie, multiple suicide attempts) or refer to the same event (ie, a single suicide attempt). The latter may occur when, for example, after making a suicide attempt, a patient has an emergency department (ED) visit followed by an inpatient hospitalization or outpatient follow-up encounters, with one (or multiple) suicide attempt codes assigned at each. In the absence of manual reviews of the narrative notes within patients’ EHRs, which cannot be performed at scale, it can be challenging to determine whether such recurrent suicide attempt codes, especially when documented over short time periods, refer to independent, distinct suicide attempts. Failure to make this important distinction can result in a form of “data leakage” in which the outcome to be predicted is included among features used for the prediction. This can result in substantial inflation of model performance [13].

To address this issue, some researchers have restricted model development to predict only the first occurrence of a suicide attempt code in a patient’s EHR [14-16]. This approach has a major limitation, however, in that a past suicide attempt is among the strongest known predictors of future suicidal behavior [17]. Thus, models that predict only the first documented suicide attempt ignore the subset of patients who may be at highest risk and thus of greatest clinical concern: those with a prior suicide attempt. Another approach is to include any previous suicide attempt codes as predictors of a subsequent suicide attempt code [18-21] thus including potential “repeat attempters” in these models. This approach, however, poses a significant risk of artificially inflating model performance if subsequent codes do not in fact refer to new suicide attempts. In other words, if a suicide attempt code instance used as an outcome actually indexes an attempt that was included a predictor, model performance will be inflated.

To minimize the risk of data leakage while retaining the option of including prior attempts as predictors, we aimed to develop an automated rule for determining whether recurrent suicide attempt codes in the EHR refer to distinct events. Such a rule might be based on relevant variables including clinical setting (eg, a suicide attempt code documented in the ED may be more likely to refer to a new suicide attempt event than one given in a non-ED setting), method (eg, suicide attempt codes that specify different methods may be more likely to refer to distinct events than codes specifying the same method), and time (eg, the more time elapsed between 2 suicide attempt codes, the less likely it may be that the codes refer to the same event). Here, we conducted a comprehensive manual EHR chart review to derive an automated rule that could identify criteria for selecting distinct suicide attempts with high confidence.

Methods
Data Source
The data source for this study was the Mass General Brigham (MGB) Research Patient Data Registry [22]. This registry covers 6.7 million patients treated in MGB-affiliated hospitals including the Massachusetts General Hospital and Brigham and Women’s Hospital in Boston.

Ethics Approval
This research was approved by the MGB institutional review board, which granted a waiver of informed consent (protocol #2018P0001508).

Case Definition and Inclusion Criteria
Details of the development of our EHR-based case definitions for suicide attempt in the MGB health care system are reported elsewhere [14,15]. In brief, we first identified candidate ICD, Ninth Revision (ICD-9) and ICD, Tenth Revision (ICD-10) codes that are likely to capture suicide attempts. Next, expert clinicians conducted manual chart reviews of 670 patients (over 3000 narrative notes) to determine a final set of codes that capture suicide attempts with a positive predictive value (PPV) of >0.70: for ICD-9, E95*, 965*, 967*, 969*, and 881*, and for ICD-10, X71*-X83*, T14.91*, T36*-T39* where the sixth character is 2 (except for T36.9, T37.9, T41.4, T42.7, T43.9, T45.9, T47.9, and T49.9, where the fifth is 2) and T51*-T65* where the sixth character is 2 (except for T51.9, T52.9, T53.9, T54.9, T56.9, T57.9, T58.0, T58.1, T58.9, T59.9, T60.9, T61.0, T61.1,
T61.9, T62.9, T63.9, T64.0, T64.8, and T65.9, where the fifth is 2).

For this study, we randomly selected a sample of 300 patients with 2 suicide attempt codes documented at least one but no more than 90 days apart (the “narrow sample”). This interval was chosen to capture codes that were given within a narrow time frame and thus potentially enriched for being “leaked” codes. In a sensitivity analysis, we randomly selected a second, smaller sample of 100 patients with 2 suicide attempt codes documented at least 1 day apart but with no other restrictions on intercode interval (the “broad sample”). A total of 31 patients appeared in both narrow and broad samples. Patients for whom we were unable to confidently locate the narrative notes corresponding to documented suicide attempt codes (eg, no narrative notes available within 30 days of the suicide attempt code date, narrative notes recorded on paper and never migrated to the EHR) were excluded after the sampling process.

**Procedure**

Under the supervision of JWS (a senior clinician with expertise in the treatment of suicidal behavior), 2 study team members (EMM and ES) manually reviewed the EHR clinical encounter data (including narrative notes) relevant to each pair of suicide attempt codes (“code pair”) per sampled patient (1015 in the narrow sample and 300 in the broad sample; 1253 unique codes across the 2 samples). Each code pair comprised a given suicide attempt code and the immediately (temporally) preceding code in a patient’s EHR. All applicable code pairs per patient were examined (including other code pairs with >90-day intervals for patients in the narrow sample). Chart reviewers assigned the following variables to each code pair: (1) whether the code pair referred to 2 distinct suicide attempts (dichotomous variable indicating distinct or not distinct suicide attempts), (2) clinical setting in which each code in the pair was documented (dichotomous variable indicating ED or non-ED [eg, outpatient and inpatient setting], (3) suicide attempt method of each code in the pair (categorical variable with 6 categories derived from previous literature: poisoning, cutting or piercing, hanging or strangulation or suffocation, jumping, firearm, and other [which included codes with no specified method]), and (4) time elapsed (in days) between codes in each pair [23]. When there were multiple encounters with suicide attempt codes on the same day, these variables were assigned to codes at the day level; see Table S1 in Multimedia Appendix 1 for an example of sampled codes (and the variables assigned to each code and code pair) for a deidentified patient.

**Data Analysis**

We defined PPV as the probability that the second code in a pair of codes identified a new suicide attempt independent of the first code in the pair. To mimic the approach that would likely be taken in building predictive models, each code pair was treated independently (ie, we did not account for the nested nature of code pairs within patients). First, for the narrow sample, we calculated (in Excel [Microsoft]) PPVs and 95% CIs by clinical setting, suicide attempt method, and intercode interval, respectively. For clinical setting, we calculated the PPVs for 4 possible code pair types: (1) both codes documented in the ED (ED/ED), (2) first code ED and second code non-ED (ED/non-ED), (3) first code non-ED and second code ED (non-ED/ED), and (4) neither code ED (non-ED/non-ED). For suicide attempt method, we calculated the PPVs of 2 possible code pair types: (1) same suicide attempt method for codes in a pair and (2) different suicide attempt methods for codes in a pair. For intercode interval, we first calculated PPVs for all 7-day intervals from 1 to 91 days, followed by collapsing across intervals from 92 days on. We then calculated the PPVs for time intervals within each of the 6 (4 clinical settings and 2 suicide attempt methods) code pair types. To derive our proposed rule, we set our benchmark PPV to 0.90. For each of the 6 code pair types, we determined the minimum time elapsed between codes (ie, interval floor) at which the PPV was at least 0.90. For a sensitivity analysis, we computed the same series of PPVs for the broad sample.

**Results**

**Descriptive Statistics**

The mean number of suicide attempt codes per patient in the narrow sample was 3.38 (SD 4.62; range 1–47). A total of 225 (75%) patients had <4 codes and 281 (93.7%) had <10 codes. A total of 210 (20.7%) code pairs had a second code reflecting a subsequent encounter for a condition for which the patient had received active treatment (indicated by a seventh “D” character).

Regarding how often the codes in a pair referred to distinct suicide attempts, of the 300 patients in the narrow sample, only 81 (27%) had more than one confirmed (by manual chart review) suicide attempt captured by the reviewed code pairs. Of the 1015 code pairs, only 180 (17.7%) referred to 2 distinct suicide attempt events. Table S1 in Multimedia Appendix 1 presents an example of sampled codes (and the variables assigned to each code and code pair) for a deidentified patient.

For clinical setting, the most common code pair types were non-ED/non-ED (n=542, 53.4%) followed by ED/ED (n=274, 27%). Regarding the 749 total non-ED codes, the most commonly represented clinical setting was inpatient (n=411, 54.9% of all non-ED codes), followed by other or unclear setting (n=149, 19.9% codes), intensive or critical care units (n=134, 17.9% codes), and outpatient (n=55, 7.3% codes). For suicide attempt method, the majority of code pairs (n=766, 75.5%) comprised 2 codes that referred to the same method. The median interval between codes in each code pair, across all codes, was 1 day. Among code pairs that referred to distinct suicide attempt events, the median interval was 35 days.

**PPVs**

**Clinical Setting**

Non-ED/ED code pairs (23 total code pairs) had the highest PPVs (0.96, 95% CI 0.87-1.04) for distinct suicide attempt events (Table 1). ED/ED pairs (274 total code pairs) had the second-highest PPVs (0.49, 95% CI 0.43-0.55). When the second code in a pair was assigned in a non-ED setting, PPVs were low (below 0.10).

In a sensitivity analysis, we excluded codes or encounters documented in inpatient settings with a prior code on the previous day from an inpatient or critical or intensive care

https://formative.jmir.org/2024/1/e46364
setting. For example, if a patient was given suicide attempt codes on three consecutive days in an inpatient setting, we only used the day 1 code. This resulted in 792 (versus 1015) analyzed code pairs. The results were overall very similar to when we did not exclude contiguous inpatient codes (Multimedia Appendix 2).

**Table 1.** Code pairs in the narrow sample defined by the clinical setting (ED\(^a\) or non-ED) of the first and second codes in each pair.

<table>
<thead>
<tr>
<th>First code clinical setting</th>
<th>Second code clinical setting</th>
<th>Code pairs (percentage of all code pairs), n (%)</th>
<th>Code pairs referring to distinct attempts, n</th>
<th>Interval between codes (days), median (Q1, Q3)</th>
<th>Interval between codes (days), mean (SD)</th>
<th>PPV(^d) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-ED</td>
<td>Non-ED</td>
<td>542 (53.4)</td>
<td>14</td>
<td>3.58 (8.64)</td>
<td></td>
<td>0.03 (0.01-0.04)</td>
</tr>
<tr>
<td>ED</td>
<td>Non-ED</td>
<td>176 (17.3)</td>
<td>10</td>
<td>6.47 (34.23)</td>
<td></td>
<td>0.06 (0.02-0.09)</td>
</tr>
<tr>
<td>Non-ED</td>
<td>ED</td>
<td>23 (2.3)</td>
<td>22</td>
<td>154.09 (286.63)</td>
<td></td>
<td>0.96 (0.87-1.04)</td>
</tr>
<tr>
<td>ED</td>
<td>ED</td>
<td>274 (27)</td>
<td>134</td>
<td>53.79 (211.77)</td>
<td></td>
<td>0.49 (0.43-0.55)</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>1015</td>
<td>180</td>
<td>21.05 (122.43)</td>
<td></td>
<td>0.18 (0.15-0.20)</td>
</tr>
</tbody>
</table>

\(^a\)ED: emergency department.  
\(^b\)Q1: first quartile.  
\(^c\)Q3: third quartile.  
\(^d\)PPV: positive predictive value.

**Suicide Attempt Method**

For suicide attempt method (same versus different method for 2 codes in a code pair), the PPVs were below 0.25 (Table 2).

**Table 2.** Code pairs defined by whether the first and second codes referred to the same or a different suicide attempt method.

<table>
<thead>
<tr>
<th>First and second code</th>
<th>Code pairs (percentage of all code pairs), n (%)</th>
<th>Code pairs referring to distinct attempts, n</th>
<th>PPV(^a) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same method</td>
<td>766 (75.5)</td>
<td>128</td>
<td>0.17 (0.14-0.19)</td>
</tr>
<tr>
<td>Different method</td>
<td>249 (24.5)</td>
<td>52</td>
<td>0.21 (0.16-0.26)</td>
</tr>
<tr>
<td>Overall</td>
<td>1015</td>
<td>180</td>
<td>0.18 (0.15-0.20)</td>
</tr>
</tbody>
</table>

\(^a\)PPV: positive predictive value.

**Intercode Interval**

Table 3 presents PPVs for code pairs broken down by 7-day (week-long) intervals; the majority (\(n=797, 78.5\%\)) of code pairs had an intercode interval of 7 days or less. The more days elapsed between 2 codes, the larger the PPV (and, fewer code pairs per strata). Table S7 in Multimedia Appendix 4 presents PPVs for code pairs broken down by interval and clinical setting (non-ED/non-ED, ED/non-ED, non-ED/ED, ED/ED), and Table S8 in Multimedia Appendix 5 presents PPVs for code pairs broken down by interval and suicide attempt method (same versus different). In another sensitivity analysis, given that ICD-9 is no longer used, we also computed all PPVs reported in Tables 1-3 when excluding code pairs with at least one ICD-9 coded event. The same pattern of findings held, with 95% CIs for all PPVs overlapping with those in Tables 1-3.
Table 3. Code pairs defined by intercode interval.

<table>
<thead>
<tr>
<th>Intercode interval</th>
<th>Code pairs (percentage of all code pairs), n (%)</th>
<th>Code pairs referring to distinct attempts, n</th>
<th>PPV(^a) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-7 days</td>
<td>797 (78.5)</td>
<td>31</td>
<td>0.04 (0.03-0.05)</td>
</tr>
<tr>
<td>8-14 days</td>
<td>48 (4.7)</td>
<td>19</td>
<td>0.40 (0.26-0.53)</td>
</tr>
<tr>
<td>15-21 days</td>
<td>31 (3)</td>
<td>17</td>
<td>0.55 (0.37-0.72)</td>
</tr>
<tr>
<td>22-28 days</td>
<td>20 (2)</td>
<td>14</td>
<td>0.70 (0.50-0.90)</td>
</tr>
<tr>
<td>29-35 days</td>
<td>17 (1.7)</td>
<td>10</td>
<td>0.59 (0.35-0.82)</td>
</tr>
<tr>
<td>36-42 days</td>
<td>17 (1.7)</td>
<td>15</td>
<td>0.88 (0.73-1.04)</td>
</tr>
<tr>
<td>43-49 days</td>
<td>10 (1)</td>
<td>8</td>
<td>0.80 (0.55-1.05)</td>
</tr>
<tr>
<td>50-56 days</td>
<td>18 (1.8)</td>
<td>14</td>
<td>0.78 (0.59-0.97)</td>
</tr>
<tr>
<td>57-63 days</td>
<td>5 (4.9)</td>
<td>3</td>
<td>0.60 (0.17-1.03)</td>
</tr>
<tr>
<td>64-70 days</td>
<td>7 (0.7)</td>
<td>6</td>
<td>0.86 (0.60-1.12)</td>
</tr>
<tr>
<td>71-77 days</td>
<td>2 (0.2)</td>
<td>2</td>
<td>1.00 (1.00-1.00)</td>
</tr>
<tr>
<td>78-84 days</td>
<td>9 (0.9)</td>
<td>9</td>
<td>1.00 (1.00-1.00)</td>
</tr>
<tr>
<td>85-91 days</td>
<td>6 (0.6)</td>
<td>6</td>
<td>1.00 (1.00-1.00)</td>
</tr>
<tr>
<td>92+ days</td>
<td>28 (2.7)</td>
<td>26</td>
<td>0.93 (0.83-1.02)</td>
</tr>
<tr>
<td>Overall</td>
<td>1015</td>
<td>180</td>
<td>0.18 (0.15-0.20)</td>
</tr>
</tbody>
</table>

\(^a\)PPV: positive predictive value.

As shown in Figure 1, across all code pairs, pairs with an interval of at least 53 days had a PPV of 0.90 (range 0.88-0.93). The interval floors meeting our benchmark PPV (at least 0.90) within each of the 6 code pair types are also labeled in Figure 1 (clinical setting) and Figure 2 (suicide attempt method). For non-ED/ED code pairs (23 code pairs), an interval floor of 1 day had a PPV of 0.96. When both codes were assigned in the ED (271 code pairs), PPV reached 0.90 when the intercode interval was at least 5 days. When the second code in a pair was documented in an ED (regardless of the setting in which the first code was documented), PPV was 0.91 when the intercode interval was 5 days (the PPV was 0.89 for 4 days). Thus, whenever the second code in a pair was documented in an ED at least 5 days after the previous code, the probability that the second code referred to an independent suicide attempt was at least 90%.

Figure 1. PPVs for interval floors by code pair types defined by clinical setting. The labeled data points indicate the interval floor at which the PPV was at least 0.90 (or the maximum PPV). Gray lines reflect PPVs for interval floors across all code pair types. Red lines refer to code pairs documented in ED (first code) and ED (second code) settings (ED/ED). Blue lines are non-ED/ED code pairs; purple lines ED/non-ED; and green lines non-ED/non-ED. ED: emergency department; PPV: positive predictive value.
Sensitivity Analysis: Broad Sample

Results from the same series of analyses in the broad sample are presented in Multimedia Appendix 6. Of the 100 patients, 45 (45%) had more than 1 confirmed suicide attempt. Of the 300 code pairs, 86 (28.7%) referred to 2 distinct suicide attempts. The median interval between codes in each pair was also 1 day. Among code pairs that referred to distinct suicide attempts, the median interval was 133 days. Overall, we found a similar pattern of PPVs (in almost all cases overlapping 95% CIs) to those from the narrow sample. Across all code pairs in the broad sample, those with an interval of at least 37 days had a PPV of 0.90 (range 0.87-0.93). When both codes were given in the ED (86 pairs), PPV reached 0.90 when the interval was at least 2 days.

Discussion

Primary Findings

Machine learning suicide risk prediction models that leverage routinely collected EHR data can outperform clinician assessment [8] and have the potential to improve how patients at risk for suicide are identified and treated. These models are typically trained using ICD codes to label suicide attempts. An under-appreciated challenge when building these models, however, is that ICD codes indexing a single suicide attempt are often used repeatedly across multiple encounters. This could create a substantial problem for models that incorporate prior suicide attempts, an established risk factor, in predicting subsequent attempts or suicidal behavior.

Some investigators side-step this issue by restricting model predictions to only the first occurrence of a suicide attempt code. This approach, however, limits the utility of prediction models by ignoring prior attempts, the best-known risk factor for suicidal behavior, and limiting their application to a subset of those at risk; prior studies indicate that nearly one-quarter of those who engage in deliberate self-harm have recurrent episodes within 3 years [24]. Here we aimed to develop a portable, automated rule for determining when recurrent suicide attempt codes refer to distinct suicide attempt events in a patient’s history. Based on chart review of clinical encounters corresponding to 1015 unique ICD code pairs, we found that, for patients with more than 1 documented suicide attempt code, repeat codes most often (>80% of the time) reflected nonindependent events, underscoring the high frequency of “leaked” suicide attempt codes. When collapsing across all clinical settings, repeat codes needed to be documented at least 53 days after the preceding code in order to refer (with probability >90%) to a new, distinct suicide attempt. However, repeat codes documented in an ED at least 5 days after the preceding suicide attempt code were likely (probability >90%) to refer to a new, distinct suicide attempt.

The most informative variables for determining whether recurrent suicide attempt codes referred to distinct suicide attempts were the clinical setting in which the codes were documented and the time elapsed between codes. First, regarding clinical setting, when a suicide attempt code was documented in an ED after the preceding code, it referred to a new suicide attempt more than half the time. Suicide attempt codes documented in non-ED settings, accounting for most of the second codes among all code pairs, however, were highly unlikely to refer to a new suicide attempt (probability <5%). This may be due to the fact that the vast majority (nearly three-quarters) of non-ED codes occurred in inpatient or intensive or critical care units, where patients may be treated over the course of several days or longer, potentially accumulating multiple suicide attempt codes that all refer to the same index event that may have prompted inpatient or intensive treatment. This pattern of findings, for one, highlights the considerable risk of treating all recurrent suicide attempt codes (especially those from non-ED settings) as distinct events, and the potential importance of using a simple rule, such as that proposed here, to identify probable distinct suicide attempt events.
Along these lines, the more time elapsed between 2 suicide attempt codes, the more likely it was the codes referred to distinct events. Combining these 2 variables—clinical setting and time elapsed—provided a simple rule for determining whether recurrent suicide attempt codes refer to distinct events with at least 90% probability. Although the accuracy of our proposed rule (at least 5 days elapsed between a code given in the ED and the preceding code) may differ in other health care systems, we recommend that others consider taking into account these 2 variables when incorporating recurrent suicide attempt codes in EHR-based suicide risk prediction models.

Perhaps surprisingly, whether the coded suicide attempt method for 2 codes in a pair was the same or different did not provide value in identifying distinct suicide attempt events. However, in the relatively small proportion of code pairs (24.5%) that referred to different methods, the most common “profile” was 1 code with a specific method (eg, poisoning and cutting or piercing) and the other code with method categorized as “other” (not a different specific method); notably, the “other” category included codes lacking any specified method. Thus, the fact that method did not help identify distinct events may largely reflect inconsistencies in how or whether the suicide attempt method is coded by providers. In contrast, neither of the other 2 variables examined (clinical setting or intercode interval) should be impacted by irregular coding practices, and thus may also be more scalable and reliable for other health care systems planning to use this or a similar rule.

Our derived rule (at least 5 days elapsed between a code from the ED and the preceding code) may have more impact on certain suicide-related prediction tasks than others. For example, it may be especially relevant when estimating patients’ risk of repeat suicidal behavior, for example after an ED visit for suicidal behavior, which could influence clinical decision-making at the point of care (eg, about discharge home or to outpatient care versus hospitalization). This rule may have less impact for other related prediction tasks, such as estimating patients’ risk of suicidal behavior after nonsuicide-related outpatient visits or broader population-based prediction efforts [25]. These results may also be less relevant for models that solely predict fatal self-harm or suicide deaths [26,27]. Future work should systematically evaluate the performance and clinical utility of models that do and do not incorporate the proposed rule for incorporating recurrent suicide attempt codes across a range of prediction goals and clinical contexts.

Our results must be considered in the context of a few key limitations. First, some of the sampled patients may have presented to hospitals outside of the MGB system for suicide attempts. In these cases, the corresponding diagnostic codes and contextual information were either unavailable or only sporadically recorded in narrative notes at subsequent clinical encounters within MGB. We also excluded sampled patients for whom chart reviewers could not confidently match data pulled from the MGB Research Patient Data Registry to the narrative notes.

Conclusions
This analysis indicates that EHR-based suicide attempt prediction models that include ICD codes for prior attempts as a predictor may be highly susceptible to bias due to data leakage in model training. Our proposed rule for circumventing this issue should minimize this bias and its inflationary effect on model performance metrics. The key variables included in our rule (clinical setting and time elapsed between codes) are widely available in health system data warehouses and should be easily integrated into EHR-based models. It is also possible that the approach taken in this study may be relevant for developing and refining machine learning models aimed to predict other episodic events of interest that can be repeatedly documented in the health record, such as unintentional overdose, domestic abuse, or episodes of violence. If effectively implemented into existing and future suicide risk prediction models, this rule could increase the robustness and validity of machine-learning based approaches to identifying the individuals at highest risk for suicide, and ultimately advance suicide prevention efforts in health care contexts on a large scale.

Acknowledgments
The authors thank the Enterprise Research Infrastructure and Services at MGB for their in-depth support and for the provision of the research patient data registry and the ERISOne Linux cluster. This work was supported by grants (NIMH R01 MH117599; JWS) and (K23MH120436; KHB) from the National Institute of Mental Health, and a gift from the Tommy Fuss Fund (JWS). JWS is a member of the Leon Levy Foundation Neuroscience Advisory Board, the Scientific Advisory Board of Sensorium Therapeutics (with equity), and has received an honorarium for an internal seminar Tempus Labs. He is principal investigator of a collaborative study of the genetics of depression and bipolar disorder sponsored by 23andMe for which 23andMe provides analysis time as in-kind support but no payments.

Data Availability
The data used in this study cannot be made publicly available due to restrictions relating to the use of EHR data.

Conflicts of Interest
None declared.
Multimedia Appendix 2
Sensitivity analysis (excluding contiguous codes from inpatient settings).

Multimedia Appendix 3
Code pairs in the narrow sample defined by specific category of suicide attempt method (poisoning, cutting or piercing, hanging or strangulation or suffocation, jumping, firearm, or other) of the first and second code in each code pair.

Multimedia Appendix 4
Code pairs in the narrow sample defined by both the clinical setting (ED or non-ED) of and the interval (in days) between the first and second codes in each pair.

Multimedia Appendix 5
Code pairs in the narrow sample defined by both suicide attempt method (same or different) and the interval (in days) between first and second codes in each pair.

Multimedia Appendix 6
Sensitivity analysis (results for broad sample).

References


**Abbreviations**

- **ED**: emergency department
- **EHR**: electronic health record
- **ICD**: International Classification of Diseases
- **MGB**: Mass General Brigham
- **PPV**: positive predictive value
Synergizing the Behavior Change Wheel and a Cocreative Approach to Design a Physical Activity Intervention for Adolescents and Young Adults With Intellectual Disabilities: Development Study

Laura Maenhout¹, MSc; Julie Latomme¹, PhD; Greet Cardon¹, PhD; Geert Crombez², PhD; Geert Van Hove³, PhD; Sofie Compernolle¹, PhD

¹Department of Movement and Sports Sciences, Ghent University, Ghent, Belgium
²Department of Experimental-Clinical and Health Psychology, Ghent University, Ghent, Belgium
³Department of Special Needs Education, Ghent University, Ghent, Belgium

Abstract

Background: There is a need for physical activity promotion interventions in adolescents and young adults with intellectual disabilities. Current interventions have shown limited effectiveness, which may be attributed to the absence of theory and a population-specific development. Combining a planning model (including theory) and cocreation with the target audience during intervention development could potentially address this gap.

Objective: This study aimed to report the systematic development of the Move it, Move ID! intervention by describing how the 8 different steps of the Behavior Change Wheel (BCW) were applied and present the results that emerged from those steps. In doing so, the (theoretical) content of the intervention is described in detail.

Methods: A total of 23 adolescents and young adults (aged 14-22 years) with mild to moderate intellectual disabilities were designated as cocreators of the intervention. Across 2 groups, 6 similar cocreation sessions were organized in each. The content and sequence of the sessions were structured to align with the 8 steps of the BCW. All sessions were recorded and transcribed verbatim. Both a deductive (ie, steps of the BCW) and inductive (ie, resonating the voice of the participants) analysis approach were applied specifically focusing on identifying and describing the findings within each of the BCW steps.

Results: After behavioral analysis (steps 1-4), 10 intervention goals were chosen and linked to Capability, Opportunity, and Motivation–Behavior components (theory within the BCW) that needed to be addressed. Psychological capability, social opportunity, and reflective motivation were emphasized as the first targets to focus on. A key finding was the urge for real-life social connectedness and social integration, which makes the social component as part of physical activity a central theme to focus on within intervention development. Judgments on the most suitable intervention functions (step 5) and behavior change techniques (step 7) were explained. When discussing the mode of delivery of the intervention (step 8), it was underscored that solely relying on a mobile health app would not fulfill participants’ social needs. Hence, the chosen intervention adopts a dyadic approach in which young individuals with intellectual disabilities are matched with peers without intellectual disabilities to engage in physical activities together, with a mobile app playing a supportive role in this partnership.

Conclusions: The transparent description of the development process highlights why certain intervention components and behavior change techniques were chosen and how they are intertwined by means of the selected intervention design. This paper provides a detailed blueprint for practitioners wanting to integrate the BCW and its associated behavior change techniques, in combination with actively involving the target group, into their intervention development for people with intellectual disabilities.

(JMIR Form Res 2024;8:e51693) doi:10.2196/51693
KEYWORDS
Behavior Change Wheel; cocreation; intervention; physical activity; adolescents; young adults; intellectual disabilities

Introduction

Physical Activity for People With Intellectual Disabilities
People with intellectual disabilities (IDs), defined as limitations in intellectual functioning (IQ of <70) and adaptive behavior emerging in childhood (age of <22 years) [1], are at higher risk of chronic health problems, such as type 2 diabetes, obesity, osteoarthritis, thyroid disorders, and cardiovascular diseases, than people without IDs [2-7]. In addition, compared with their peers without disabilities, individuals with IDs have less access to health care services, face increased polypharmacy, have higher poverty rates, encounter social isolation, and engage more in behaviors that put their health at risk (eg, unhealthy nutrition and physical inactivity) [4].

Promoting physical activity (PA) may be one avenue to reduce increased health problems. PA has been shown to have beneficial effects on the physical and psychosocial health of people with IDs [8,9]. Nevertheless, they are less physically active than their peers without IDs [10-16]. The World Health Organization (WHO) recommends that adolescents engage in at least 60 minutes of moderate to vigorous PA (MVPA) per day and participate in muscle- and bone-strengthening activities 3 days per week. Adults are recommended to perform at least 150 minutes of moderate-intensity PA or 75 minutes of vigorous-intensity PA throughout the week or an equivalent combination of moderate- and vigorous-intensity activity supplemented by performing activities twice a week to strengthen muscles and bones [17]. Since 2020, these global PA guidelines include groups such as people living with IDs.

It is a positive trend that, for the first time, there is attention to the specific target group in the WHO PA guidelines. However, it should be noted that the evidence is primarily based on individuals without IDs, and some argue that disability-specific guidelines are necessary [2].

A 2016 systematic review including 15 studies described that only 9% of adults with IDs achieved minimum PA guidelines (with a range of 0%-46%), measured using both objective and self-reported measurement tools [18]. Different PA guidelines were used as outcome measures in the included studies, such as 150 minutes of MVPA per week (in bouts of >10 minutes), 30 minutes of MVPA for at least 5 days per week, 20 minutes of mild exercise ≥4 times per week, 12 bouts of MVPA in 4 weeks (retrospectively), and >10,000 steps per day. A systematic review the year after (2017) reported that, in 5 out of 17 studies that assessed MVPA through accelerometry in participants with IDs (aged 6-72 years), none of the participants met the PA guidelines of 150 minutes of MVPA per week and adults and 60 minutes of MVPA per day for children and adolescents. In the remaining 12 studies, the percentage of participants with IDs who met the guidelines ranged from 6% to 66% (mostly because of the use of different protocols to measure PA) [19].

Both reviews concluded that only a small number of individuals with IDs meet the PA guidelines, indicating that they are less active than the general population [18,19].

Current PA Interventions and Their Effectiveness
Although PA research in people with IDs has been growing over the last decade, this field has been underresearched. A PA Series in The Lancet (2021) stated that, between 1999 and 2019, <5% of all articles in the 5 highest-impact medical journals focused on people with disabilities (not limited to IDs) and <7% of these addressed PA or health [2]. A systematic review from 2019 on the effectiveness of PA interventions among participants with IDs of all ages identified only three effective randomized controlled trials out of 9 [20]: (1) a 10-week progressive resistance training program in adolescents and young adults (aged 14-22 years) with Down syndrome in Australia [21], (2) a 12- to 16-month multicomponent diet and PA program in adults (aged 20-66 years) with mild to moderate IDs in Sweden [22], and (3) an 8-month PA and fitness program in “fast-walking” older adults with mild to moderate IDs (aged >40 years) in the Netherlands [23]. The success of these randomized controlled trials was mainly attributed to the following factors: (1) practical support from others (eg, a mentor) in guiding and helping participants with IDs through the intervention and for increasing motivation, (2) establishing a routine that involves regular PA as well as the adaptability of an intervention to the specific routines of both carers and participants, (3) the simplicity of an intervention by adapting interventions to the specific needs of the participants, and (4) familiarity with the intervention [20]. None of the 9 interventions in the systematic review by Hassan et al [20] included a technological component (ie, eHealth or mobile health [mHealth]). However, there seems to be no reason why digital interventions would not be feasible in this target group [24]. In the study by Ptomey et al [25], 95% of the participants, aged 14.9 (SD 2.2) years on average, reported that using a tablet computer was easy and enjoyable. It is then no surprise that, in recent years, there has been a growing interest in the development of digital interventions for individuals with IDs [26-30].

The number of effective PA interventions for people with IDs remains limited. A potential reason for the limited effectiveness is currently attributed to the lack of a theoretical framework for intervention development and the difficulty in concretizing behavior change techniques (BCTs) in an understandable way for this population [10,20,31,32]. A 2017 systematic review on the use of BCTs in lifestyle change interventions for people with IDs, for example, concluded that 73% of the studies aiming to improve PA in the target group did not use any theoretical framework [31]. Nevertheless, the use of a theoretical framework is an important prerequisite for intervention effectiveness [2,20,33-36]. Furthermore, when examining theory-based interventions for people with IDs, concerns have been raised regarding the suitability of the theories used (eg, social cognitive theory, theory of planned behavior, and self-determination theory) as a starting point for designing interventions for this specific target group. These theories may
not sufficiently address the specific challenges faced by people with IDs. More specifically, these theories tend to be specific and detailed, yet they may not encompass the complete spectrum of potential influences on behavior within this particular target group and often concentrate on individual-level factors [20,31,32].

**Applying the Behavior Change Wheel and a Cocreative Approach to Build Theory-Based PA Interventions**

The Behavior Change Wheel (BCW) is a planning model aimed at guiding a scientific and systematic intervention development process [33,37]. The BCW contains a behavioral theory at its heart, the Capability, Opportunity, and Motivation–Behavior (COM-B) model, which encompasses the full range of influences contributing to the behavior of interest [33,38]. A total of 3 behavioral components are summarized in the COM-B model, which states that, for each behavior to occur, individuals need capability (physical and psychological), opportunity (physical and social), and motivation (reflective and automatic) [33,37,39].

The COM-B model is in turn linked to the Theoretical Domains Framework (TDF) [33,40], which subdivides the COM-B model into 14 domains. The BCW further formulates 9 intervention functions linked to 93 BCTs [41] and 7 policy types with the aim of modifying each of the 3 COM-B components and, thus, changing behavior. The COM-B model describes the minimal factors that behavioral scientists agree on to achieve behavior change and has been developed with interdisciplinary research in mind [37]. It is an open model and relatively easy to communicate, especially with vulnerable groups. In recent years, this model has demonstrated applicability in the context of PA among people with IDs and their carers [32,42]. This study chose the BCW as a planning model for intervention development because of its practical use and feasibility in combination with a cocreative approach. Current lifestyle modification approaches for this target group lack a robust foundation addressing their unique needs [10,32,43]. Therefore, deeply engaging with this group and customizing approaches to promote their PA is vital. Unfortunately, individuals with IDs are seldom heard in research, and interviews with caregivers often take precedence, potentially overshadowing their authentic experiences [43]. Neglecting the perspectives of individuals with IDs can undermine intervention acceptability, comprehensibility, and feasibility [10,44-48]. To clarify, previous intervention studies have reported that some BCTs may be too complex for the target group (eg, self-monitoring through the use of pedometers) [31,49]. The cocreative approach (in combination with the BCW planning model) in this study will aid in determining which BCTs might be most appropriate for people with IDs or adapting them if necessary through collaboration.

**Aims**

This paper aimed to (1) report the systematic development of the Move it, Move ID! intervention by describing how the different steps of the BCW were applied and (2) present the results that emerged from those steps. In doing so, we described the (theoretical) content of the Move it, Move ID! intervention in detail.

**Methods**

**Participants and Recruitment**

It was prioritized to focus on young people with IDs as cocreators rather than their parents or teachers because of the historical pattern of marginalization in previous research on intervention development [43]. Through purpose sampling, 2 class groups of adolescents or young adults with mild to moderate IDs aged between 13 and 22 years (ie, age of special needs secondary education in Flanders, Belgium) were recruited to participate in the cocreation sessions. In February 2021 and March 2021, a total of 2 physical education (PE) teachers from different special needs schools in Flanders were contacted to explain the purpose and design of the project via email and phone. They were asked whether they were interested in involving one of their classes in cocreating a PA promotion intervention. Each PE teacher subsequently suggested 1 class group to take part. All adolescents from the selected classes (classes A and B) were invited to participate during the first visit, in which written informed consent from all participants and passive consent from their parents were obtained (Table 1).

Class A comprised 14 adolescents aged between 17 and 22 years with a mild to moderate level of ID (mean age 20.33, SD 1.94 years; 3/14, 21% female). Class B comprised 9 adolescents aged between 14 and 15 years with mild IDs (mean age 14.22, SD 0.44 years; 6/9, 67% girls). This aligned with cocreation guidelines, which recommend groups of 10 to 12 cocreators [45,50]. A detailed description of (the recruitment of) participants, as well as the ethical process (next subsection), can be found in the study by Maenhout et al [51].
Table 1. Merging the Behavior Change Wheel (BCW) with a cocreational approach.

<table>
<thead>
<tr>
<th>BCW phase</th>
<th>Researchers’ tasks</th>
<th>Cocreation part with participants with IDs$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stage 1: understand the behavior</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1: define the health problem in behavioral terms</td>
<td>Determine the health problem in behavioral terms using the literature: insufficient PA$^b$ in people with IDs</td>
<td>No input was gathered from the participants with IDs in the first 2 steps as we relied on the literature to define the health problem and select the target behavior. Furthermore, the PI$^c$ is currently affiliated with the Department of Movement and Sports Sciences (Ghent University), which is why we focused on PA.</td>
</tr>
<tr>
<td>Step 2: select the target behavior</td>
<td>Select the target behavior: increasing PA levels in adolescents and young adults with IDs</td>
<td>No input was gathered from the participants with IDs in the first 2 steps as we relied on the literature to define the health problem and select the target behavior. Furthermore, the PI$^c$ is currently affiliated with the Department of Movement and Sports Sciences (Ghent University), which is why we focused on PA.</td>
</tr>
<tr>
<td>Step 3: specify the target behavior and formulate intervention goals</td>
<td>Specify the target behavior by:</td>
<td>Cocreation session 1:</td>
</tr>
<tr>
<td></td>
<td>• Generating a nonexhaustive list of all potential barriers and facilitators that may be relevant to the target behavior</td>
<td>• Introduction session (ie, explanation of the project and its purpose, process of informed consent, and getting to know each other)</td>
</tr>
<tr>
<td></td>
<td>• Describing these barriers and facilitators as what needs to be targeted in the intervention (who needs to do it, what do they need to do differently to achieve change, where and when do they need to do it, and how often and with whom do they need to do it)</td>
<td>Cocreation session 2:</td>
</tr>
<tr>
<td></td>
<td>• Formulating 10 intervention goals based on the ranking by the cocreators</td>
<td>• Comapping barriers to and facilitators of PA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• What PAs are they currently performing? What do they like or dislike?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cocreation session 3:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Explore the most important barriers and facilitators on which the intervention should focus by voting and ranking them by importance</td>
</tr>
<tr>
<td>Step 4: link intervention goals to COM-B$^d$ components and TDF$^e$ domains</td>
<td>Select the components of the COM-B model and the theoretical domains of the TDF for each intervention goal</td>
<td>No input was gathered from the participants with IDs in this step as their input (from step 3) only needed to be linked to the theoretical components of the COM-B and TDF.</td>
</tr>
<tr>
<td><strong>Stage 2: identify intervention options</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 5: select intervention functions</td>
<td>Select intervention functions using the APEASE$^f$ criteria from the BCW guide [33]</td>
<td>No input was gathered from the participants with IDs in this step as intervention functions were first considered to be feasible by the project team. However, an open-minded perspective was adopted in which only intervention functions that were deemed not feasible by the project team were removed.</td>
</tr>
<tr>
<td>Step 6: identify policy categories</td>
<td>Not applied as designers limited to a specific policy lever are directed immediately to step 7 [33]</td>
<td>N/A$^g$</td>
</tr>
<tr>
<td><strong>Stage 3: identify content and implementation options</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 7: identify BCTs$^h$</td>
<td>Choose the most appropriate BCT(s) based on the following:</td>
<td>Cocreation session 4:</td>
</tr>
<tr>
<td></td>
<td>• The BCW guide [33]</td>
<td>• Select BCTs and identify whether selected BCTs would suit the target group or how they can be redesigned to work for them</td>
</tr>
<tr>
<td></td>
<td>• Input from participants with IDs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• APEASE criteria (expert consultation)</td>
<td></td>
</tr>
</tbody>
</table>
Cocreation part with participants with IDs

- Cocreation session 5:
  - Explore facilitators of and barriers to mHealth
  - How can we make an mHealth intervention as feasible and acceptable as possible for them?

- Cocreation session 6—this session was no longer about intervention development but about the study itself, such as the following:
  - Explore the opinion of participants with IDs on the best recruitment strategy
  - Find out which incentives they would prefer

| BCW | Researchers’ tasks | Cocreation part with participants with IDs
|-----|--------------------|-----------------------------------------------|
| Step 8: identify mode of delivery | • Choose mode of delivery based on the following:  
  • The literature (ie, high potential of using an mHealth intervention)  
  • Input from participants with IDs | • Cocreation session 5:  
  • Explore facilitators of and barriers to mHealth  
  • How can we make an mHealth intervention as feasible and acceptable as possible for them? |

*ID: intellectual disability.
*PA: physical activity.
*PI: principal investigator.
*COM-B: Capability, Opportunity, and Motivation–Behavior.
*TDF: Theoretical Domains Framework.
*APEASE: Affordability, Practicality, Effectiveness and Cost-Effectiveness, Acceptability, Side Effects or Safety, and Equity.
*N/A: not applicable.
*BCT: behavior change technique.
*mHealth: mobile health.

Ethical Considerations

All participants and their parents or legal guardians received detailed and accessible information regarding the study design and purpose as well as data handling. To ensure privacy, the data were pseudonymized and only accessible to the researchers or their appointed representatives. Data confidentiality was always maintained. In consultation with the data protection officer of Ghent University (Belgium), the legal basis was changed from “active informed consent” of parents or legal guardians to “public interest,” although this did not exempt researchers from informing participants. This meant that parents or legal guardians needed to contact the researchers only if they disagreed with their child’s participation and, thus, researchers did not require active consent from parents or guardians to commence. The participants with IDs themselves were required to provide their active consent, which is why the first session involved a thorough, step-by-step review of the information and consent process with time for discussion. Young people with IDs participated voluntarily and could withdraw at any time. In appreciation of their participation, all participants received 2 cinema tickets, about which they were informed when they decided to participate. This study received approval from the Ethical Committee of the Faculty of Psychology and Educational Sciences at Ghent University, Belgium (2021/056 LM).

Combining the BCW Development Process and a Cocreational Approach

Overview

From April 2021 to June 2021, the 8 steps of the BCW were systematically followed for intervention development (Figure 1) [33]. In parallel, input was gathered from adolescents and young adults with IDs through 6 cocreation sessions (Table 1). The entire process was a mix of theoretical underpinnings (ie, COM-B), the domain expertise of the researcher, and the lived experiences of the target group (ie, cocreation sessions). The 6 sessions took place in their classrooms, each during 2 consecutive class hours. For a comprehensive explanation of the cocreation process, the methods used, and the participants’ experiences, we refer interested readers to our previously published paper [51].
**Stage 1: Understand the Behavior**

In step 1, the literature was reviewed by the principal investigator (PI; LM) to articulate the (health) problem in behavioral terms, this being “insufficient PA in people with IDs.” The next steps were then to select (step 2) and specify (step 3) the target behavior of the intervention, this being “increasing PA levels in adolescents and young adults with IDs,” by defining who needs to do it, what needs to be done differently to achieve change, where and when they need to do it, and how often and with whom they need to do it. This was done by generating a nonexhaustive list of possible barriers to and facilitators of PA for adolescents and young adults with IDs based on both the literature and information gathered in the second cocreation session (Table 1). Owing to the cocreative approach, insights from the literature (brought in by the PI) and input from the target group were intertwined (eg, visual cards of barriers and facilitators were developed by the PI inspired by the literature, which were brought up when the participants themselves could not come up with barriers and facilitators anymore [51]). In cocreation session 3, these barriers and facilitators were ranked according to importance by the target group. The most important barriers and facilitators were described as what needed to be targeted in the intervention and, consequently, formulated as the intervention goals. In step 4, these intervention goals were then assigned by the PI to the specific components of the COM-B model and theoretical domains of the TDF. No direct cocreation session was organized within this step as their input (from step 3) only needed to be linked to the theoretical components of the COM-B and TDF. However, this does not deviate from the essence of cocreation as the PI established these connections based on all the input provided by the participants.

**Stage 2: Identify Intervention Options**

In step 5, the BCW guide links COM-B components and TDF domains to 9 intervention functions [33]. Consequently, the broader research group of the PI (ie, the Physical Activity and Health research group) held expert meetings to decide which intervention functions were most suitable to work with based on the Affordability, Practicability, Effectiveness or Cost-Effectiveness, Acceptability, Side Effects or Safety, and Equity criteria [33]. These criteria are recommended by the BCW guide to make strategic judgments on the most appropriate intervention functions. No direct input from the cocreators was sought in this case, either. However, we approached this step with an open-minded perspective and only removed the intervention functions that were deemed not feasible by the project team. All other intervention functions were retained, allowing the cocreators to continue shaping the direction of development. The sixth step was to consider which policies would support the delivery of the intervention functions identified in step 5 [33]. However, as the researchers within this project did not have access to policy levers, step 6 was not applied. This is also described in the BCW guide by stating that “designers limited to a specific policy lever are directed to step 7 to identify BCTs” [33].

**Stage 3: Identify Content and Implementation Options**

In step 7, the BCW guide proposes the most appropriate BCTs for each intervention function (selected in step 5) [33]. In each of these, a distinction is made between “BCTs used most frequently and less frequently” [33]. For feasibility reasons, we focused primarily on the most frequently used BCTs during the development process. However, for the fourth cocreation session on BCTs, we also explored the less frequently used BCTs and selected relevant ones based on our expertise with the target group. The aim of this cocreation session was to find out which BCTs were understandable and feasible for adolescents and young adults with IDs and how BCTs could be adapted to meet these criteria. On the basis of the Affordability, Practicability, Effectiveness or Cost-Effectiveness, Acceptability, Side Effects or Safety, and Equity criteria and on input from the cocreators with IDs, a decision was made on which BCTs to include in the
Move it, Move ID! intervention. Finally, the eighth step was to identify the best way to deliver the intervention (ie, mode of delivery). As research has shown that the use of technology (ie, mHealth) is feasible and has high potential in adolescents and young adults with IDs [24,25,52], the target group was asked in the fifth cocreation session about their preferences and barriers to and facilitators of mHealth use.

Analysis
All the cocreation sessions were recorded and transcribed verbatim. A combination of a deductive (ie, 8 steps of the BCW) and inductive (ie, resonating the voice of the participants) analysis approach was applied specifically focusing on identifying and describing the findings from each of the steps of the BCW.

Results

Stage 1: Understand the Behavior

Step 1: Define the Health Problem in Behavioral Terms
Few people with IDs are sufficiently physically active [18,19].

Step 2: Select the Target Behavior
An increase in the total volume of PA should be targeted rather than aiming to meet the WHO guidelines regarding MVPA as even small positive changes in PA levels are associated with health benefits among people with IDs [53].

Step 3: Specify the Target Behavior and Formulate Intervention Goals
Multimedia Appendix 1 [32,54-63] provides an overview of 72 barriers to and 66 facilitators of PA for adolescents and young adults with IDs based on (1) a review of the literature by the PI in preparation for the cocreation sessions and (2) input from cocreators with IDs during these sessions. The appendix is divided into intrapersonal, interpersonal, and contextual factors, reflecting the multifaceted and complex nature of the influences on PA in this population. In the third cocreation session, participants ranked the barriers and facilitators according to their importance, providing guidance on which ones to address in the intervention. The 10 most important barriers (in the opposite direction, these would be facilitators) were identified: (1) the need for social connectedness, (2) the lack of practical support within the PA context, (3) the absence of a role model, (4) the need for others around them who also engage in PA, (5) the lack of confidence in their own abilities and body image, (6) the need for knowledge about the (health) benefits of PA, (7) the lack of knowledge about the different PA options available, (8) the low motivation to engage in PA, (9) the difficulty in setting goals, and (10) the need for help to incorporate PA into their existing schedules (ie, goal conflict) as they often depend on others for this. Evidently, this top list does not mean that the other barriers and facilitators were not relevant for some individuals at particular times, but in view of feasibility, it was decided to prioritize and primarily address those that were identified as the most important.

Previous studies have proposed schools as the ideal setting for PA promotion [8,32,64]. Participants with IDs in this study indicated that they are sufficiently encouraged at school to engage in PA via compulsory PE classes. However, they expressed difficulties in being physically active during leisure time. In the cocreation sessions, they expressed a preference for an intervention during their leisure time (ie, at home or in the community setting) rather than a school-based intervention:

I think it’s best to go somewhere else. Then you have something separate from school. That you are really away. When you come back to school, that you can start again with a fresh head. [Cocreator 1; cocreation session 3; group A]

 […] that you just keep your activities outside school and that you don’t keep it here between these four walls. [Cocreator 2; cocreation session 3; group A]

Textbox 1 summarizes the specifics of the target behavior gathered during the first 3 steps: who will perform the behavior; what needs to be done differently; and when, where, how often, and with whom it needs to be done.

Finally, the PI formulated 10 intervention goals targeting the most important barriers chosen by the cocreators (Table 2).
Specify the target behavior (step 3 of the Behavior Change Wheel).

<table>
<thead>
<tr>
<th>Who needs to perform the behavior?</th>
<th>Flemish adolescents and young adults aged between 14 and 22 years with mild to moderate intellectual disabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>What does the person need to do differently to achieve the desired change?</td>
<td>Address the 10 most important barriers or facilitators (described in Table 2)</td>
</tr>
<tr>
<td>When do they need to do it?</td>
<td>During leisure time (weekdays+weekends)</td>
</tr>
<tr>
<td>Where do they need to do it?</td>
<td>In the community setting or at home</td>
</tr>
<tr>
<td>How often do they need to do it?</td>
<td>Not specified</td>
</tr>
<tr>
<td>With whom do they need to do it?</td>
<td>Together with someone (at this stage, it was not specified yet who this someone could be, but the need for social connectedness during physical activity did emerge as the main barrier or facilitator in both groups)</td>
</tr>
</tbody>
</table>
### Table 2: Linking of intervention goals to Capability, Opportunity, and Motivation–Behavior (COM-B) components and Theoretical Domains Framework (TDF) domains (step 4 of the Behavior Change Wheel).

<table>
<thead>
<tr>
<th>COM-B component and relevant TDF domain</th>
<th>Most important barriers or facilitators</th>
<th>Intervention goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capability</td>
<td>Psychological</td>
<td></td>
</tr>
<tr>
<td>Psychological</td>
<td>Knowledge</td>
<td></td>
</tr>
<tr>
<td>• Insufficient knowledge about options for PA(^a), where, what suits the person best, and what are the barriers and how to counter them</td>
<td>Adolescents and young adults with IDs(^b) need a better understanding of where, when, and how to engage in PA; they need to be offered a range or variety of PA options they can choose from.</td>
<td></td>
</tr>
<tr>
<td>• Adolescents and young adults with IDs need a better understanding of their (own) barriers to PA and how to counter them</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral regulation</td>
<td>Difficulty in setting up PA goals (mostly because of a lack of knowledge about PA options)</td>
<td>Adolescents and young adults with IDs need to be facilitated/supported in formulating specific PA goals.</td>
</tr>
<tr>
<td>• Difficulty with planning PA (eg, mostly because of the dependency on others and goal conflict)</td>
<td>Adolescents and young adults with IDs need to be facilitated/supported in planning PA.</td>
<td></td>
</tr>
<tr>
<td>Opportunity</td>
<td>Social</td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>Social influences</td>
<td></td>
</tr>
<tr>
<td>• Lack of social connectedness; having no one to do PA with (eg, friends or loved ones)</td>
<td>Adolescents and young adults with IDs need to have the opportunity to engage in PA together with someone (ie, social connectedness).</td>
<td></td>
</tr>
<tr>
<td>• Not having a role model (ie, seeing other people engage in PA as well)</td>
<td>Adolescents and young adults with IDs need a role model regarding PA.</td>
<td></td>
</tr>
<tr>
<td>• No guidance during PA or no practical support</td>
<td>Adolescents and young adults with IDs need to have more (social and practical) support from others when engaging in PA.</td>
<td></td>
</tr>
<tr>
<td>Motivation</td>
<td>Reflective</td>
<td></td>
</tr>
<tr>
<td>Reflective</td>
<td>Intentions</td>
<td></td>
</tr>
<tr>
<td>• No motivation to engage in PA</td>
<td>Adolescents and young adults with IDs need to be encouraged in feeling a sense of enjoyment when engaging in PA.</td>
<td></td>
</tr>
<tr>
<td>Beliefs about capabilities</td>
<td>Insecure about own capabilities and skills (eg, afraid of doing something wrong, afraid of the reaction of others, afraid of PA being too difficult, or afraid of being laughed at)</td>
<td>Adolescents and young adults with IDs need to build self-confidence regarding PA (ie, increase self-image and confidence in their ability to perform certain activities).</td>
</tr>
<tr>
<td>• Insecure or ashamed about weight or body shape</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beliefs about consequences</td>
<td>Lack of awareness about the health consequences of physical inactivity</td>
<td>Adolescents and young adults with IDs need a better understanding of the benefits of PA.</td>
</tr>
</tbody>
</table>

\(^a\)PA: physical activity.

\(^b\)ID: intellectual disability.

### Step 4: Link Intervention Goals to COM-B Components and TDF Domains

In step 4, the PI assigned the 10 intervention goals to the specific COM-B and TDF components of the BCW (Table 2). From this behavioral analysis, it can be inferred that psychological capability, social opportunity, and reflective motivation would be the first targets to focus on for increasing PA levels in adolescents and young adults with IDs.

### Stage 2: Identify Intervention Options and Steps 5 and 6 (Identify Intervention Functions and Policy Categories)

Linking the selected COM-B components and TDF domains from step 4 to the intervention functions proposed by the BCW guide, all 9 intervention functions could be applied. In total, 7 intervention functions were chosen to further focus on: education, persuasion, incentivization, training, environmental restructuring, modeling, and enablement (see the detailed argumentation in Multimedia Appendix 2). The 2 other intervention functions were not selected as they were deemed (1) less practicable to apply as a research team (ie, restriction).
and (2) less acceptable or unlikely to have an impact on adolescents or young adults with IDs (ie, coercion).

**Stage 3: Identify Content and Implementation Options**

**Step 7: Identify BCTs**

**Overview**

A total of 12 BCTs were selected to proceed with. We have outlined our selection and reasoning for each selected BCT within the specific intervention function in the following sections. *Multimedia Appendix 3* provides a detailed explanation of all the BCTs that were considered for the 7 intervention functions that came out of step 5, along with the accompanying rationale for why they were chosen and others were not.

**Education**

Participants with IDs expressed a lack of knowledge about PA options (eg, what is out there, what suits the person best, and where can it be done). For this reason, it was considered valuable to provide information on various PA options. However, the only BCTs formulated within the taxonomy by Michie et al [41] related to providing information are pertaining to consequences (ie, social, emotional, environmental, and health). Although participants mentioned the value of information about the health benefits of PA in previous stages, we collectively decided not to place a direct emphasis on information provision within our intervention. Participants do not desire an intervention centered on “learning” or “teaching” (see also their preference for an intervention outside the school context). According to them, the focus should be on enjoyment. Nonetheless, we anticipate that the target audience may indirectly experience positive effects through the intervention. In this regard, the BCT “salience of consequences” (under the persuasion intervention function) seemed more applicable as it focuses on using methods to specifically emphasize the consequences of performing a behavior, making them more memorable, which goes beyond mere information provision about these consequences. “Feedback on behavior” was also selected as participants indicated that they would like to receive feedback on how well they are performing the behavior.

**Persuasion**

The BCT “credible source” was valued by participants, but opinions varied on its presentation. Some preferred health professionals using fun visual communication, whereas others liked animated movies. In this target group, experts or influencers explaining PA benefits in an engaging way were considered more appealing than scientific videos. Furthermore, the significance of “verbal persuasion about capability” was strongly emphasized. Given the low self-efficacy within this population, offering verbal persuasion to counteract self-doubt was deemed highly valuable for adolescents/young adults with IDs.

**Incentivization**

Owing to the prominent role of social factors, we observed that the BCT “social reward” would be highly motivating for this target group.

**Environmental Restructuring**

Cocreators highlighted the importance of social connectedness and support in encouraging PA. As a result, we expect the greatest impact from recognizing and meeting their social needs, which entails a “restructuring of their social environment.”

**Modeling**

Participants expressed that it would be motivating to witness others engaging in PA around them, whether in person or indirectly through influencers such as on TikTok (serving as role models). The cocreators showed enthusiasm for involving influencers they admired to encourage PA. Considering budget limitations, it would not be feasible for us to incorporate a well-known influencer into the intervention. However, this does indicate that “demonstration of the behavior” might be an interesting BCT to include.

**Enablement**

The entire development process highlighted a strong emphasis on the importance of social support and social connections, whether from friends or individuals with expertise in PA (ie, social support BCT—practical, emotional, and unspecified). Participants expressed increased confidence when they could openly discuss their goals and challenges with friends, and their motivation to engage in PA was significantly higher when they could do it with others rather than exercising alone. Peer support was generally preferred, although younger adolescents with IDs (aged 14 years) also mentioned the potential for support from family members. Furthermore, participants agreed that having a list of goals to choose from would make it easier for them rather than having to come up with their own goals (ie, goal setting BCT). Most participants recognized the importance of “action planning” as a valuable BCT. However, insights from teachers revealed that adolescents and young adults with IDs often struggle with tasks such as maintaining a personal agenda or planner, which is typically managed by parents or carers. Therefore, it would be crucial to offer guidance during action planning. Creating a detailed action plan independently, including specifics such as what, when, where, and with whom, seemed challenging and burdensome for this group. Simplicity and minimizing cognitive effort were emphasized as essential factors. Similarly, the collaborative review and adjustment of the behavioral goal with individuals with IDs based on their progress was seen as advantageous. It was considered feasible to engage in close negotiation with them to either retain the same goal, make minor adjustments, or establish a new goal if necessary (referred to as “reviewing behavior goals”). The primary focus in this case is on shared decision-making and active involvement.

**Training**

The only BCT that we considered including under the “training” intervention function is “demonstration of the behavior.” However, we view this as more related to modeling behavior rather than as actual behavior demonstration within a training context. In the course of our intervention development, it became evident that the primary focus should be on addressing social needs and creating enjoyable experiences rather than on formal training in activities. Therefore, the “training” intervention function was omitted from this phase onward.
Step 8: Identify Mode of Delivery

On the basis of the literature, an mHealth app appeared to be a good and feasible approach for adolescents and young adults with IDs and, therefore, was verified during the fifth cocreation session. The cocreators indicated that they preferred an mHealth app with a straightforward design that clearly indicated its purpose and functionality (e.g., through an introductory video). They suggested that the app should be visually appealing, with minimal text, bright colors, and no foreign-language words. They also suggested that a game component or chat feature would be of added value. Cocreators would not use an app that they had to pay for, was childish, or looked rather old-fashioned. They mentioned preferring not to receive too many notifications (i.e., no more than 1 notification per day). Finally, this is a group that often faces negative comments and experiences of failure. When talking about mHealth, this also emerged as an aspect to be considered (e.g., by keeping the reactions that can be given to each other in an app controlled).

At the end of the fifth cocreation session, cocreators indicated that an app alone would not be sufficient to encourage them to engage in (more) PA. They suggested that an app could be integrated into a broader intervention but not be a stand-alone intervention. More specifically, the desire for social connection with peers and social integration in real life was found to be a more important theme in intervention development. Therefore, the decision was made to focus on a buddy system as many people with IDs reported a lack of friendships with peers outside school, resulting in decreased opportunities to engage in PA during leisure time. To facilitate this buddy partnership, we chose to work with a buddy without IDs who could offer practical support during the intervention period, which reduced the reliance on context alone (i.e., parents or carers) to guide the intervention implementation.

Move It, Move ID! Intervention

On the basis of the systematic steps of the BCW combined with a cocreational approach, the Move it, Move ID! intervention ultimately consists of a buddy partnership with a supporting app (i.e., dyadic intervention). Figure 2 illustrates the development process, showing how COM-B components, intervention functions, and BCTs are intertwined with the selected intervention design. A more in-depth description can also be found in Multimedia Appendix 3.

Figure 2. Visualization of the development process showing how Capability, Opportunity, and Motivation–Behavior (COM-B) components; intervention functions; and behavior change techniques (BCTs) are intertwined with the selected intervention design. PA: physical activity; TDF: Theoretical Domains Framework.

During an intervention period of 3 months, adolescents and young adults with IDs will be paired with a buddy without IDs of the same age range and encouraged to try out weekly PAs in Ghent (Flanders, Belgium). Buddies without IDs will be students (aged 17-23 years) of the coauthors of this paper and will receive 3 short training sessions (i.e., maximum of 1 hour per session) on their role and responsibilities as a buddy.

Although the buddy partnership forms the core of the intervention, a supporting app will also be provided in which buddies and participants with IDs will be in direct contact with each other (an explanation of the scope and screenshots of the supporting app can be found in Multimedia Appendix 4). The app is considered a private space between participants with IDs and their buddies without adding parents or carers to the app. The PI will add a range of activities (e.g., walking a shelter dog, dancing, playing Kubb, and undertaking an altitude trail) to the app at the start of the intervention. Participants will have the autonomy to choose whether they want to try an activity by agreeing (swiping right) or disagreeing (swiping left) with a proposed activity. When both the participant with IDs and the
buddy agree with a certain proposed activity, they will receive a pop-up to a chat function to make arrangements and schedule this activity on their shared agenda. The buddy will take the lead in this process. During an activity, the buddy can provide feedback such as how well they perform the behavior or words of encouragement. On the app pinboard, pairs can share photos of the activity they performed together, give comments, and also rate the activity afterward. This allows them to keep track of successful activities and identify less enjoyable ones.

Discussion

Principal Findings

This paper describes the systematic, theory-driven development of a lifestyle intervention to promote PA in adolescents and young adults with IDs using the BCW planning model combined with cocreation sessions involving the target group. The purpose of this transparent and detailed description was 2-fold. First, it aimed to develop a PA promotion intervention by identifying intervention components and BCTs that address the specific needs of this target group. Second, it aimed to encourage future researchers and intervention developers interested in PA among adolescents and young adults with IDs to apply a theoretical planning model in combination with cocreation when designing similar interventions or take the insights described into account in their own intervention development. By transparently describing the theory and BCTs that underpin the intervention, researchers are facilitated in broader evaluations to explore their driving mechanisms. In doing so, we adhered to the Medical Research Council guidelines, which emphasize the importance of theorizing how an intervention works and what works in which setting and identifying its other impacts [65]. This discussion will first delve deeper into the key findings regarding the development of the Move it, Move ID! intervention followed by a reflection on the experience of the development process by combining the BCW and cocreation.

The development process underscored the essential importance of collaborating with the target group as the intervention looks different from what the research team had envisioned in the project proposal (ie, develop an mHealth app for young people with IDs targeted at promoting PA) with a thorough argument that adaptations are necessary from the perspective of the target group itself. In that regard, the combination of actively involving the target audience and applying a clear and scientific planning model was crucial. The most prominent planning models that are currently proposed to guide the development of effective interventions are Intervention Mapping [72] and the BCW [33]. Intervention Mapping includes 6 different steps to rigorously select determinants, performance, and change objectives using appropriate methods and strategies [72]. Although Intervention Mapping is comprehensive, its level of detail makes it more complex and, thus, less feasible, especially in combination with cocreation [73]. The BCW, in contrast, is more open, practical, and flexible as it was developed with interdisciplinary application in mind [37]. However, in applying the BCW within this project, it was noticed that its openness and flexibility could also lead to variable interpretations, with judgments from the researchers often required throughout the development process (eg, step 5). The variations in intervention development mainly depended on the resources available to the project team (eg, affordability and practicability). Moreover, even within this small research team of the Move it, Move ID! intervention, different steps within the BCW were sometimes interpreted differently. Some researchers saw the formulation of barriers to and facilitators of PA as belonging to steps 2 and 3 (as it was described in this paper), whereas others ascribed this to step 4 [74]. In our opinion, assigning these aspects to a certain step will not differ much from the behavioral diagnosis one will eventually arrive at. We consider it more important to discuss the different steps thoroughly within the research team so that the decisions made are well informed and can be argued for. By going through the different steps of the BCW, we learned that interventions can look different depending on the choices made without necessarily making one intervention better than the other. Further research should subsequently indicate which
Interventions prove to be effective and why (i.e., identifying driving mechanisms [65]). This could potentially lead to the formulation of guidelines outlining the best possible choices that could be made during intervention development within a specific target group and setting.

Nevertheless, by applying the theoretical planning model, the PI had a clear goal in mind in setting up the structure and flow of the cocreation sessions. In doing so, the BCW was instrumental in identifying an informed behavioral diagnosis and choosing which BCTs would be most applicable to have an impact on PA behavior change within this target group and setting. Although the literature suggests that the use of theory in intervention development is key [2,20,33-36], a 2019 meta-analysis formulated that the effectiveness of interventions would be less influenced by whether they are theoretically developed than by the specific BCTs used [75]. In contrast, we believe that both (i.e., theory and choice of BCTs) are intertwined. A 2017 systematic review found that lifestyle change interventions for people with IDs aimed at improving PA levels typically used 5.9 BCTs, with “provide information on consequences of behavior in general,” “plan social support/social change,” “provide instruction on how to perform the behavior,” and “goal setting (behavior)” being the most frequently used BCTs [31]. However, 73% of the studies did not use any theoretical framework for intervention development [31]. After completing the full behavioral diagnosis based on the BCW, we included 12 BCTs in our intervention. This is not to say that the inclusion of more BCTs would be better but, rather, that the transparent description of the BCW steps made more evident why these specific BCTs were chosen and how they are intertwined by means of the intervention design. This demonstrates why we believe that the use of theory and the selection of BCTs are strongly connected.

Linking cocreation to the BCW, our goal was to create an intervention that starts with the experiences of the target group. This approach was intended to enhance the effectiveness and sustainability of the intervention by making it more suitable and acceptable for the target audience [10,44-48]. Cocreation with the target audience extended well beyond the described cocreation sessions for intervention development in this project. As the project progressed toward the effect study, ongoing collaboration continued with 2 coresearchers with IDs (i.e., inclusive research [48]). These coresearchers maintained regular meetings (every 2 weeks) with the PI (LM) at the Department of Movement and Sports Sciences (Ghent University), actively engaging in various facets of the project. Their responsibilities included assessing prototypes of the app; offering feedback on the training of buddies; testing measurement instruments (comprising questionnaires, interviews, and accelerometers); providing insights into the recruitment strategy; contributing to the development of promotional materials such as flyers, information letters, and informed consent forms; and participating in efforts to enhance the project’s visibility among their peers, classmates, and other stakeholders. This ongoing collaboration with the coresearchers was purposefully designed to ensure the continued accessibility of the project even beyond the initial phase of intervention blueprinting. To conclude, the described intervention development addresses an important and often overlooked population that experiences health disparities and is at higher risk of physical inactivity and related health issues. This study highlights the importance of considering the unique requirements of people with IDs to develop tailored interventions that effectively meet their needs.

Limitations and Strengths

This study has some limitations. First, a wide age range of adolescents and young adults with IDs was included, which might make us question whether this intervention is applicable to both an individual aged 13 years and one aged 22 years. Indeed, younger adolescents with IDs (i.e., aged 14 and 15 years) did indicate that they would be open to involving parents as buddies within an intervention, whereas this was not the case for young adults (i.e., aged 17-22 years old). Choosing a tighter age limit (e.g., ages of 13-16 years or 17-22 years) is recommended in future intervention development. Second, of the 23 cocreators, 5 (22%) had a comorbidity with autism spectrum disorder, and 1 (4%) adolescent had attention-deficit/hyperactivity disorder. This is considered a limitation as previous research has found different effects on PA among youth who have IDs and youth who have other developmental disabilities in addition to IDs [8], suggesting that further comparison of PA experiences between these groups is warranted. Within the further intervention development, little weight was given to these comorbidities. In contrast, we can also conclude that their perspective was included from the start of intervention development as they also acted as cocreators and this was not an exclusion criterion. Third, following the prioritization of young people with IDs as cocreators in the initial stages of blueprinting an intervention idea, we were unable to gather input from buddies (peers intended to be matched with the participants with IDs) and consider the broader context of individuals with IDs in the actual development phase of the intervention. This constraint was due to the project’s timeline. In light of this constraint, we recommend that future intervention developers consider including these stakeholders in subsequent phases of intervention development. Their perspectives and insights are invaluable in creating interventions that are comprehensive, inclusive, and truly reflective of the needs and dynamics of the entire participant group. The greatest strength of this study was the fact that a theoretical planning model was used in combination with cocreation to develop a PA promotion intervention for this target group. In this way, it addressed the two main reasons why current interventions often prove to be ineffective: (1) a lack of use of theory and (2) a lack of population-specific research. To the best of our knowledge, this is the first study that describes the collaborative development of a PA promotion intervention for and with adolescents and young adults with IDs. Within the Move it, Move ID! project, the decision was made to work only with participants with mild to moderate IDs; consequently, the findings cannot be extended to the target group of severe or profound IDs. Although future research should focus on the representation of all people with IDs in health research, the fact that a specific group was chosen to truly tailor an intervention to their needs can also be seen as a strength.
Conclusions
The Move it, Move ID! intervention was developed based on the BCW in combination with cocreation. Going through this process was seen as an added value by the research team, which makes it highly recommended to allocate adequate time, budget, and experienced scientific staff for intervention development. By systematically identifying the needs of young people with IDs and linking them to theoretical concepts step by step, cocreators with IDs emphasized the importance of face-to-face interactions and social components in PA promotion interventions. They indicated that relying solely on an mHealth app would not fulfill their social needs. The intervention will consist of a dyadic approach in which young individuals with IDs are paired with a peer without IDs to engage in PAs together, with an app solely providing support within this partnership. The detailed and transparent development process described is a valuable blueprint for practitioners wanting to integrate the BCW and its associated BCTs, in combination with actively involving the target group, into their intervention development for people with IDs.

Acknowledgments
The authors wish to thank the adolescents or young adults and teachers who expressed their opinions during intervention development for this project. This work was supported by the Research Foundation–Flanders under grant 11F3621N (2020-2024).

Data Availability
The data sets generated and analyzed during this study are not publicly available to protect participants’ privacy and confidentiality because of the small number of participants but are available from the corresponding author upon reasonable request. In addition, the upcoming effect study has been preregistered at the Open Science Framework, where all other materials concerning this project can be found.

Authors' Contributions
LM, SC, G Cardon, G Crombez, and GVH conceptualized the study. LM collected the data (ie, organized the cocreation sessions) and wrote the original draft. JL assisted in combining the Behavior Change Wheel planning model and a cocreational approach. SC, JL, G Cardon, G Crombez, and GVH edited the manuscript and provided feedback. All authors have read and approved the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Barriers to and facilitators of physical activity for adolescents or young adults with intellectual disabilities.
[PDF File (Adobe PDF File), 170 KB - formative_v8i1e51693_app1.pdf ]

Multimedia Appendix 2
Selection of intervention functions using the Affordability, Practicality, Effectiveness and Cost-Effectiveness, Acceptability, Side Effects or Safety, and Equity criteria.
[PDF File (Adobe PDF File), 201 KB - formative_v8i1e51693_app2.pdf ]

Multimedia Appendix 3
Selection and reasoning for each selected and nonselected behavior change technique.
[PDF File (Adobe PDF File), 194 KB - formative_v8i1e51693_app3.pdf ]

Multimedia Appendix 4
Scope of the Move it, Move ID! app.
[PDF File (Adobe PDF File), 525 KB - formative_v8i1e51693_app4.pdf ]

References


Abbreviations

- BCT: behavior change technique
- BCW: Behavior Change Wheel
- COM-B: Capability, Opportunity, and Motivation–Behavior
- ID: intellectual disability
- mHealth: mobile health
- MVPA: moderate to vigorous physical activity
- PA: physical activity
- PE: physical education
- PI: principal investigator
- TDF: Theoretical Domains Framework
- WHO: World Health Organization
Assessing and Improving Data Integrity in Web-Based Surveys: Comparison of Fraud Detection Systems in a COVID-19 Study

Stephen Bonett¹, RN, PhD; Willey Lin¹, MB; Patrina Sexton Topper¹, BSN, MS, PhD; James Wolfe¹, MS; Jesse Golinkoff², MPH; Aayushi Deshpande², MPhil; Antonia Villarruel¹, RN, PhD; José Bauermeister¹, MPH, PhD

¹School of Nursing, University of Pennsylvania, Philadelphia, PA, United States
²Department of Psychology, Ashoka University, Sonepat, India

Corresponding Author:
Stephen Bonett, RN, PhD
School of Nursing
University of Pennsylvania
418 Curie Boulevard
Philadelphia, PA, 19104
United States
Phone: 1 2155734299
Email: stepdo@nursing.upenn.edu

Abstract

Background: Web-based surveys increase access to study participation and improve opportunities to reach diverse populations. However, web-based surveys are vulnerable to data quality threats, including fraudulent entries from automated bots and duplicative submissions. Widely used proprietary tools to identify fraud offer little transparency about the methods used, effectiveness, or representativeness of resulting data sets. Robust, reproducible, and context-specific methods of accurately detecting fraudulent responses are needed to ensure integrity and maximize the value of web-based survey research.

Objective: This study aims to describe a multilayered fraud detection system implemented in a large web-based survey about COVID-19 attitudes, beliefs, and behaviors; examine the agreement between this fraud detection system and a proprietary fraud detection system; and compare the resulting study samples from each of the 2 fraud detection methods.

Methods: The PhillyCEAL Common Survey is a cross-sectional web-based survey that remotely enrolled residents ages 13 years and older to assess how the COVID-19 pandemic impacted individuals, neighborhoods, and communities in Philadelphia, Pennsylvania. Two fraud detection methods are described and compared: (1) a multilayer fraud detection strategy developed by the research team that combined automated validation of response data and real-time verification of study entries by study personnel and (2) the proprietary fraud detection system used by the Qualtrics (Qualtrics) survey platform. Descriptive statistics were computed for the full sample and for responses classified as valid by 2 different fraud detection methods, and classification tables were created to assess agreement between the methods. The impact of fraud detection methods on the distribution of vaccine confidence by racial or ethnic group was assessed.

Results: Of 7950 completed surveys, our multilayer fraud detection system identified 3228 (40.60%) cases as valid, while the Qualtrics fraud detection system identified 4389 (55.21%) cases as valid. The 2 methods showed only “fair” or “minimal” agreement in their classifications ($\kappa=0.25$; 95% CI 0.23-0.27). The choice of fraud detection method impacted the distribution of vaccine confidence by racial or ethnic group.

Conclusions: The selection of a fraud detection method can affect the study’s sample composition. The findings of this study, while not conclusive, suggest that a multilayered approach to fraud detection that includes conservative use of automated fraud detection and integration of human review of entries tailored to the study’s specific context and its participants may be warranted for future survey research.

(JMIR Form Res 2024;8:e47091) doi:10.2196/47091

KEYWORDS
web-based survey; data quality; fraud; survey methodology; COVID-19; survey; fraud detection; Philadelphia; data privacy; data protection; privacy; security; data; information security; data validation; cross-sectional; web-based
**Introduction**

Web-based survey research has become increasingly common in recent years, particularly because of its ability to reach broad populations efficiently and economically [1]. Web-based surveys involve inviting potential respondents to complete questionnaires through digital platforms that manage how questions are presented and how data are collected and stored [2,3]. These research methods have been used in response to the difficulties faced in traditional survey methods (i.e., recruiting participants using flyers, newspaper or radio or television advertisements or spreading by word of mouth and collecting data in person using computer-assisted survey instruments or over the phone), especially in reaching underrepresented populations [4,5]. The advantages of web-based surveys include eliminating the requirement for face-to-face interaction, offering flexible access to surveys, removing transportation and logistical barriers, and preserving anonymity. In recent years, COVID-19 pandemic restrictions limited opportunities for in-person research and provided additional justification for researchers to adopt web-based study designs while leveraging social media recruitment methods to reach diverse populations [6-9].

As web-based recruitment and survey methods in health research have become more ubiquitous and refined, so too have methods of web-based research fraud [10,11]. Fraud can manifest in multiple ways. For example, individuals may misrepresent themselves in order to appear eligible for a study or may submit duplicate surveys in order to receive multiple incentive payments. Additionally, fraudulent data may also come from automated operations enacting fraud at a large scale, often referred to as “bots” [11,12]. These methods are often used to target surveys offering participation compensation payments and can be lucrative when aimed at large web-based surveys, even those offering small payments [9,13]. Such fraud poses risks not only to research resources but also, importantly, to the integrity of research findings, as fraudulent data can distort results and undermine data quality. Specifically, fraudulent responses can introduce additional random noise or potentially add systematic bias to the data [14-16].

In response, researchers, companies operating in the digital research space (e.g., Qualtrics) [17], and organizations interested in digital data integrity (e.g., Google) [18] have developed methods to address fraudulent activity. The research community has crafted recommendations for fraudulent data identification and participant identity verification protocols [13,19,20]. Platforms specializing in web-based survey research such as Qualtrics [17] and Amazon Mechanical Turk [21] have also developed fraud detection features that accompany their services. While these proprietary systems for fraud detection offer a simple, automated approach to improving data quality, little information is available about the mechanisms they use [22]. Fraud detection systems often obscure details about how their validation process functions as an important strategy to protect the integrity of the fraud detection system, making it more difficult for fraudulent participants to circumvent protections. However, obfuscation also introduces questions about how fraud detection algorithms alter study samples and whether they introduce bias into analyses [20].

Little research has compared how fraud detection strategies impact study sample composition or examined their comparative effectiveness in correctly identifying fraud [22-25]. By accurately identifying and removing fraudulent responses to web-based surveys, research can improve data quality and strengthen the overall rigor of their methods. Robust, reproducible, and context-specific methods of accurately detecting fraudulent responses are needed to ensure integrity and maximize the value of web-based survey research. This paper aims to (1) describe the multilayer fraud detection techniques we developed and implemented in a large web-based survey collecting data about attitudes, beliefs, and behaviors related to COVID-19; (2) examine the degree of agreement between our multilayer fraud detection strategy and the proprietary fraud detection system used by Qualtrics; and (3) compare the study samples that resulted when using each of the 2 fraud detection methods.

**Methods**

**Study Design**

We collected data from November 2021 through February 2022 for the PhillyCEAL Common Survey, a cross-sectional study using a web-based survey to assess how the COVID-19 pandemic and response have impacted individuals, neighborhoods, and communities across the city of Philadelphia, Pennsylvania. The Checklist for Reporting Results on Internet E-surveys was used to guide the reporting of our methods and results (Multimedia Appendix 1) [26]. The Qualtrics web-based survey platform was used to design the survey and automatically capture responses in a database. The usability and technical functionality of the survey were tested by the study team before launching the survey. Individuals were eligible to participate if they (1) resided within Philadelphia County (coterminal with the city limits) and (2) were at least 13 years of age. We recruited participants through advertisements on social media platforms (e.g., Facebook, Instagram, Twitter, and Reddit) and referrals from community partners (including one partner that provided study recruitment materials to individuals via door-to-door canvassing). The recruitment process directed individuals to a voluntary, open web-based survey, where they completed a screener consent form and answered a series of questions to determine eligibility and record basic demographic information. We did not allow participants to change their answers through a back button feature.

Eligible participants were automatically directed to the full study consent form. Following consent, participants were asked to complete a 20-minute questionnaire about their experiences, behaviors, and beliefs about COVID-19, risk of infection, testing, vaccination, treatment, and knowledge and beliefs about COVID-19 clinical trials. Participants completed 1 of 3 slight variations of the survey (i.e., the adult survey, the parent survey, and the youth survey), where additional questions or slight changes to wording were used on the parent and youth surveys. Participants completed only 1 of these 3 potential variations (i.e., survey groups were mutually exclusive), with participants of any age who reported having minor children completing the parent survey (103 questions), participants ages 25 years or
younger and not having minor children completing the youth survey (126 questions), and all other participants completing the adult survey (92 questions). To reduce participant burden, we used adaptive questioning to reduce the number of questions displayed based on their answers. Participants saw an average of 6 questions per page.

Participants were excluded if they did not complete the entire survey (n=2930) or did not provide a residential zip code matching 1 of the 48 zip codes of Philadelphia County (n=647). Participants confirmed eligible and marked as complete by both Qualtrics automation and our manual review of the data were compensated with a US $15 electronic gift card. To protect participant data, the web survey data were downloaded to a secure university server, deidentified by replacing contact information with unique IDs, and stored in a restricted folder with password protection.

**Ethical Considerations**

All participants completed an informed consent process before proceeding to the survey. Ethics approval was obtained from the institutional review board at the University of Pennsylvania (protocol 848650).

**Preliminary Fraud Protection**

Given the prevalence of fraud and duplicate responses in web-based survey research, we used several strategies to prevent fraudulent participants from accessing and completing the survey, serving as a baseline level of fraud protection for both our multilayer and the Qualtrics detection methods. Our preliminary line of defense against fraud was targeted toward nonhuman interferences such as bots. To proceed with the survey, all respondents had to pass a built-in Turing test provided by Qualtrics using Google’s reCAPTCHA (version 2) antifraud technology [27]. Depending on the respondent’s on-device saved data such as browser cookies, they either had to click a checkbox or solve a simple image challenge to pass the reCAPTCHA (version 2) test.

Since sophisticated bots can trick the reCAPTCHA (version 2) test [28], we added a honeypot question as a second line of defense against bots. Honeypots are survey questions hidden from rendering on the screen using custom JavaScript code [11]. They are, therefore, invisible to human respondents but accessible to bots that do not rely on what is rendered on screen. Since the honeypot is not visible to human respondents, any responses to the honeypot would immediately disqualify the entire survey response and end the survey.

We also created unique URLs for each recruitment source and advertisement campaign for the study. The unique URLs enabled us to identify the origin (recruitment source and ad campaign) of each survey response. This allowed us to individually monitor and suspend links that became the target of fraudulent survey responses.

**Multilayer Fraud Detection Methods**

**Real-Time Exclusion of Fraudulent Responses**

We implemented a suite near real-time data verification procedures to assess the veracity of data as they were collected, including techniques based on recommendations from prior research as well as several manual checks that were developed specifically for this study. A research team member would individually inspect submitted responses and label responses as fraudulent using the following criteria: (1) participants were asked to provide their residential address and the colloquial name of their neighborhood. Responses were marked as fraud if the neighborhood name provided did not match a standard Philadelphia neighborhood name corresponding to the residential address provided or any adjacent neighborhood; (2) the residential address provided did not match an existing address in Philadelphia County [20,23,29]; (3) the survey had the same start times and stop times plus or minus 1 minute as 2 or more other submitted surveys (rapid survey submission) [11,29]; (4) the respondent’s email address matched a previously enrolled participant’s email address [11,20]; (5) the zip code provided as part of the residential address was nonstandard (ie, a post office box code or a unique code) [20,23]; (6) the residential address had already been reported by at least 2 other respondents [20,23]; or (7) the URL from which the response was referred did not match any of the URL links distributed by the study team during recruitment. Responses labeled as fraud during real-time validation were not eligible to receive survey compensation.

**Automated Post Hoc Identification of Fraudulent Responses**

We developed a set of automated post hoc techniques designed to detect fraud that our real-time procedures may not have captured. Three criteria were developed for this post hoc fraud identification based on recommendations from prior research [11,20,22,23,29-32]. Since these criteria only identify suspicious entries and do not definitively prove that an entry is fraudulent, responses were labeled as fraud only if they satisfied 2 or more criteria. We settled on using a threshold of 2 criteria (rather than 1 or 3) in order to balance concerns about the potential of each of our 3 criteria to incorrectly label a participant as fraud with the necessity to exclude causes that showed strong evidence of fraud. The criteria were as follows:

1. In response to a free text item at the end of the survey soliciting additional comments or questions from the participant, the submitted survey included text that was identical to text submitted by other respondents. We considered a free text response an identical match if it was among free text entries of 1 word or greater that were repeated 100 or more times, free text entries of 2 words or greater that were repeated 10 or more times, or free text entries of 3 words or greater that were repeated 3 or more times (see Table S1 in Multimedia Appendix 2 for list of unique text strings excluded and their frequency in the full set of responses).
2. The IP address of a response belonged to a virtual private network or data center or originated from outside the United States, as determined by using a security service for proxy and virtual private network detection and IP location information [33]; and
3. Responses provided in the main survey were inconsistent with responses to the same items in the screener for one or more key items that would not be expected to be variable
(ie, age, zip code, number of adults living at home with the participant, number of minors living at home with the participant, Hispanic or Latinx ethnicity, “Have you ever been tested for COVID-19?,” and “Have you received at least 1 dose of the COVID-19 vaccine?”).

Qualtrics Fraud Detection Methods
Qualtrics is a widely used web-based survey platform that allows users to create surveys with complex flow logic and customizable visual design. Qualtrics surveys are easily optimized for use on mobile devices and can display a wide variety of question types on both computer and mobile phone interfaces. Another key strength of the Qualtrics platform is its integration of 1-click translation, allowing users to quickly switch between various languages. This was crucial for our study, which recruited participants from diverse populations across Philadelphia and was available in English, Spanish, and Mandarin. In addition to these valuable features, Qualtrics also offers tools for detecting fraudulent survey responses. This automated and user-friendly system for fraud detection has the potential to help researchers improve data quality in their web-based surveys. Given the lack of research exploring how these consumer tools compare to existing published protocols for fraud detection, we sought to compare our multilayer fraud detection methods to the system used by Qualtrics.

The Qualtrics fraud detection system relies on Google’s reCAPTCHA (version 3) and Imperium’s RelevantID antifraud technologies. Both tools rely on proprietary machine learning models that analyze passive and behavioral data, browser interactions, and respondent metadata to identify abuse and fraud [18,34,35]. Unlike the reCAPTCHA (version 2) test respondents had to solve at the start of the survey, bot detection using reCAPTCHA (version 3) does not present respondents with an image challenge nor block respondents and bots from proceeding with the survey. Instead, it returns a score (Q_ReCaptchaScore) between 0.0 and 1.0 that Qualtrics records as part of the survey response. We used the recommended 0.5 score as the threshold for fraud, where a score under 0.5 is deemed likely to be a bot [17,18].

Like reCAPTCHA (version 3), RelevantID does not prevent bots from completing the survey. Instead, it attaches a score (Q_RelevantIDFraudScore) between 0 and 130 to each survey response. We followed Qualtrics’ recommendation in interpreting a score ≥30 as fraudulent and likely a bot [17]. In addition to bot detection, RelevantID identifies duplicate responses through digital fingerprinting and proprietary detection algorithms [34]. Qualtrics then attaches another score (Q_RelevantIDDuplicateScore) between 0 and 100 to the survey response. We followed the suggested score threshold where any score ≥75 is considered a duplicate [17].

Statistical Analysis
Agreement and Comparative Performance
The classification tables were created to display the degree of agreement between the 2 fraud detection methods for the full sample and for each of the 3 survey-type categories (ie, adult, parent, and youth).

Impact of Fraud Detection Method on Sample Characteristics
Descriptive statistics were computed for the full sample of responses, the subset classified as valid by our multilayer fraud detection method, and the subset classified as valid by the Qualtrics fraud detection method. As these 3 sets of responses are not mutually exclusive, we did not directly compare them statistically.

To test for differences between fraudulent and valid responses as classified by each fraud detection method, statistical comparisons were conducted for key study variables between the mutually exclusive sets of responses classified as fraudulent or valid within each method. Specifically, we used chi-square tests for categorical variables, 2-tailed t-tests for normally distributed continuous variables, and Mann-Whitney U tests for continuous variables that were not normally distributed. The results of these analyses are presented in Table S2 in Multimedia Appendix 2 for the multilayer method and Table S3 in Multimedia Appendix 2 for Qualtrics.

To assess the degree to which the 2 fraud detection methods would impact the distribution of a key study variable, the point estimate and 95% CI were calculated for vaccine confidence by racial or ethnic group for each fraud detection method and for the entire sample without any fraud mitigation.

Variations in Survey Responses During Study Period
A time-series plot was created to show the cumulative responses to the study survey over time and their fraud classification by each of the 2 fraud detection methods. This plot highlights the periods in which social media recruitment campaigns are active and can also shed light on how the 2 fraud detection methods diverge in their classification of responses during different periods of high survey response. Additionally, we present a time-series plot showing the proportion of responses classified as fraud across the study period, including smooth locally weighted smoothing lines to visualize the trends over time. All analyses were performed with R (version 4.1.0; R Foundation for Statistical Computing).

Results
Multilayer Fraud Detection Methods
A total of 7950 completed survey responses were received. See Figures 1 and 2 for an overview of fraud detection results from our multilayer fraud detection methods. Using the real-time exclusion criteria of the multilayer fraud detection method, 4207 (52.92%) entries were classified as fraud. Of those classified as fraud, 1242 (29.52%) reported a neighborhood name that did not match their residential address, 648 (15.4%) provided an invalid residential address, 1397 (33.21%) displayed rapid survey submission, 42 (1%) used a repeated email address, 77 (1.83%) reported a nonstandard zip code, 398 (9.46%) reported a residential address that was used more than twice, and 403 (9.58%) did not have a valid recruitment URL. After the real-time exclusion, 3743 (47.08%) cases remained classified as valid.
Our automated post hoc fraud detection criteria identified additional cases as fraud. Of the remaining 3743 initially valid cases, 1561 (41.70%) cases had a duplicate response in the free text entry item, 394 (10.53%) cases had an IP address from outside the United States or from a virtual private network, and 619 (16.54%) had inconsistencies between the screener and main survey on at least 1 key item. Using our “2-strike” rule, we classified an additional 515 (13.76%) responses as fraud for meeting at least 2 of the above criteria. Thus, our multilayer fraud detection strategy classified a total of 4722 (59.40%) entries as fraud and 3228 (40.60%) entries as valid.

**Qualtrics Fraud Detection Methods**

The Qualtrics fraud detection methods identified 498 (6.26%) cases as fraud by the RelevantID FraudScore, and 938 (11.80%) cases as duplicates by the RelevantID DuplicateScore. The Qualtrics fraud detection strategy classified a total of 3561 (44.79%) entries as fraud (i.e., meeting one or more of the 3 criteria above) and 4389 (55.21%) entries as valid.

**Agreement and Comparative Performance**

Table 1 presents confusion matrices showing the degree of agreement between our multilayer fraud detection method and the Qualtrics fraud detection method for the full sample and each survey-type category. The intrarater reliability indicated “fair” or “minimal” agreement between the 2 methods for the full sample ($\kappa=0.25$; 95% CI 0.23-0.27), “moderate” or “weak” agreement for the adult ($\kappa=0.48$; 95% CI 0.43-0.53) and youth ($\kappa=0.50$; 95% CI 0.43-0.58) surveys, and “slight” or “none” agreement for the parent survey ($\kappa=0.13$; 95% CI 0.10-0.15) [36,37].
### Table 1. Confusion matrix and interrater reliability (κ) between our multilayer fraud detection system and the Qualtrics fraud detection system for the full sample, only adult surveys, only parent surveys, and only youth surveys.

<table>
<thead>
<tr>
<th></th>
<th>Fraud (multilayer)</th>
<th>Valid (multilayer)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full sample</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid (Qualtrics)</td>
<td>2627(^a)</td>
<td>934(^c)</td>
</tr>
<tr>
<td>Fraud (Qualtrics)</td>
<td>2095(^c)</td>
<td>2294</td>
</tr>
<tr>
<td><strong>Adult survey</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid (Qualtrics)</td>
<td>299</td>
<td>166(^c)</td>
</tr>
<tr>
<td>Fraud (Qualtrics)</td>
<td>174(^c)</td>
<td>904</td>
</tr>
<tr>
<td><strong>Parent survey</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid (Qualtrics)</td>
<td>2184</td>
<td>710(^c)</td>
</tr>
<tr>
<td>Fraud (Qualtrics)</td>
<td>1848(^c)</td>
<td>1102</td>
</tr>
<tr>
<td><strong>Youth survey</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid (Qualtrics)</td>
<td>144</td>
<td>58(^c)</td>
</tr>
<tr>
<td>Fraud (Qualtrics)</td>
<td>73(^c)</td>
<td>288</td>
</tr>
</tbody>
</table>

\(^a\)κ=0.25; 95% CI 0.23-0.27.  
\(^b\)Values in italics represent agreement between the 2 methods.  
\(^c\)Values represent disagreement between the 2 methods.  
\(^d\)κ=0.48; 95% CI 0.43-0.53.  
\(^e\)κ=0.13; 95% CI 0.10-0.15.  
\(^f\)κ=0.50; 95% CI 0.43-0.58.

We conducted sensitivity analyses to assess the impact of choosing a “2-strike rule” for our post hoc fraud detection rather than a “1-strike rule” or a “3-strike rule.” Compared to the “2-strike rule,” which resulted in 515 additional cases being classified as fraud during the post hoc phase of fraud detection, the “1-strike rule” would have classified 2047 additional cases as fraud, and the “3-strike rule” would have classified 12 additional cases as fraud. In terms of agreement with Qualtrics’ fraud detection methods, the “1-strike rule” would have resulted in a κ of 0.20 (95% CI 0.19-0.22) for the full sample, and the “3-strike rule” would have resulted in a κ of 0.24 (95% CI 0.22-0.26) for the full sample.

Additionally, we explored how the 2 fraud detection strategies compared in their ability to classify cases with validated email addresses as valid entries. Validated email addresses were defined as email addresses ending in “.edu” or “.gov,” indicating an institutional affiliation. Of the 168 cases with validated emails, the multilayer fraud detection system correctly classified 166 (98.81%) as valid, while the Qualtrics fraud detection system correctly classified only 126 (75%) as valid.

### Impact of Fraud Detection Method on Sample Characteristics

Decisions about which fraud detection strategies to use can impact the results of web-based survey research. Table 2 presents the descriptive statistics for sociodemographic variables, survey metric variables, and key study outcome variables on 3 versions of the data set: the full data set with no fraud detection (n=7950), the cases identified as valid by our multilayer fraud detection methods (n=3228), and the cases identified as valid by the Qualtrics fraud detection methods (n=4389). As these sets are not mutually exclusive, we cannot compare them directly; however, there are clear differences in the distributions of many study variables between the 3 sets. When comparing entries classified as fraud to those classified as valid for each of the 2 fraud detection methods (ie, mutually exclusive sets), all study variables, except for lifetime COVID-19 testing for the multilayer fraud detection, were found to be significantly different for both methods (Tables S2 and S3 in Multimedia Appendix 2).

Table 3 showcases in detail how a key variable of interest to researchers may be affected by using different fraud detection methods. In this data set, vaccine confidence among White respondents was greater when using our multilayer fraud detection (µ=0.867; 95% CI 0.851-0.882) when compared to Qualtrics fraud detection (µ=0.782; 95% CI 0.766-0.798). A similar pattern is seen for Hispanic or Latinx respondents and Black or African American respondents.
Table 2. Demographics, survey metrics, and key study responses in overall sample, multilayer valid set, and Qualtrics valid set.

<table>
<thead>
<tr>
<th></th>
<th>Full sample (N=7950)</th>
<th>Multilayer valid set (n=3228)</th>
<th>Qualtrics valid set (n=4389)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>35.54 (9.70)</td>
<td>38.09 (12.15)</td>
<td>37.01 (10.81)</td>
</tr>
<tr>
<td><strong>Race or ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latinx</td>
<td>1188 (14.9)</td>
<td>254 (7.9)</td>
<td>571 (13)</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>135 (1.7)</td>
<td>7 (0.2)</td>
<td>34 (0.8)</td>
</tr>
<tr>
<td>Asian</td>
<td>311 (3.9)</td>
<td>219 (6.8)</td>
<td>221 (5)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>1856 (23.3)</td>
<td>728 (22.6)</td>
<td>853 (19.4)</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>42 (0.5)</td>
<td>11 (0.3)</td>
<td>10 (0.2)</td>
</tr>
<tr>
<td>White</td>
<td>4272 (53.7)</td>
<td>1889 (58.5)</td>
<td>2600 (59.2)</td>
</tr>
<tr>
<td>Multiracial or others</td>
<td>146 (1.8)</td>
<td>120 (3.7)</td>
<td>100 (2.3)</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Woman</td>
<td>4253 (53.5)</td>
<td>2028 (62.8)</td>
<td>2645 (60.3)</td>
</tr>
<tr>
<td>Man</td>
<td>3571 (44.9)</td>
<td>1108 (34.3)</td>
<td>1663 (37.9)</td>
</tr>
<tr>
<td>Transgender or gender diverse</td>
<td>105 (1.3)</td>
<td>76 (2.4)</td>
<td>64 (1.5)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>21 (0.3)</td>
<td>16 (0.5)</td>
<td>17 (0.4)</td>
</tr>
<tr>
<td><strong>Sexual orientation, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bisexual</td>
<td>361 (4.5)</td>
<td>242 (7.5)</td>
<td>262 (6)</td>
</tr>
<tr>
<td>Gay</td>
<td>231 (2.9)</td>
<td>101 (3.1)</td>
<td>147 (3.3)</td>
</tr>
<tr>
<td>Lesbian</td>
<td>142 (1.8)</td>
<td>69 (2.1)</td>
<td>65 (1.5)</td>
</tr>
<tr>
<td>Straight (ie, not gay, lesbian, or bisexual)</td>
<td>7039 (88.5)</td>
<td>2682 (83.1)</td>
<td>3791 (86.4)</td>
</tr>
<tr>
<td>Others</td>
<td>94 (1.2)</td>
<td>84 (2.6)</td>
<td>71 (1.6)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>83 (1)</td>
<td>50 (1.5)</td>
<td>53 (1.2)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than high school</td>
<td>215 (2.7)</td>
<td>49 (1.5)</td>
<td>65 (1.5)</td>
</tr>
<tr>
<td>High school or equivalent</td>
<td>1013 (12.7)</td>
<td>299 (9.3)</td>
<td>479 (10.9)</td>
</tr>
<tr>
<td>Some college</td>
<td>1890 (23.8)</td>
<td>579 (17.9)</td>
<td>964 (22)</td>
</tr>
<tr>
<td>College graduate</td>
<td>3935 (49.5)</td>
<td>1672 (51.8)</td>
<td>2253 (51.3)</td>
</tr>
<tr>
<td>Graduate degree</td>
<td>882 (11.1)</td>
<td>620 (19.2)</td>
<td>617 (14.1)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>15 (0.2)</td>
<td>9 (0.3)</td>
<td>11 (0.3)</td>
</tr>
<tr>
<td><strong>Survey type, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adult</td>
<td>1543 (19.4)</td>
<td>1070 (33.1)</td>
<td>1078 (24.6)</td>
</tr>
<tr>
<td>Parent</td>
<td>5844 (73.5)</td>
<td>1812 (56.1)</td>
<td>2950 (67.2)</td>
</tr>
<tr>
<td>Youth</td>
<td>563 (7.1)</td>
<td>346 (10.7)</td>
<td>361 (8.2)</td>
</tr>
<tr>
<td><strong>Survey metrics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survey duration (minutes), median (IQR)</td>
<td>23.46 (18.38-38.10)</td>
<td>22.02 (18.13-32.57)</td>
<td>22.82 (18.52-35.13)</td>
</tr>
<tr>
<td>User language=Spanish, n (%)</td>
<td>127 (1.6)</td>
<td>22 (0.7)</td>
<td>53 (1.2)</td>
</tr>
<tr>
<td><strong>Key study variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ever tested for COVID-19=yes, n (%)</td>
<td>6968 (87.6)</td>
<td>2840 (88)</td>
<td>3903 (88.9)</td>
</tr>
<tr>
<td>Ever COVID-19–positive, n (%)</td>
<td>5836 (83)</td>
<td>2498 (87.3)</td>
<td>3294 (83.8)</td>
</tr>
</tbody>
</table>
### Table 3. COVID-19 vaccine confidence (somewhat confident or very confident) grouped by race compared across the 2 fraud detection methods.

<table>
<thead>
<tr>
<th>Race or ethnicity</th>
<th>Multilayer valid set (n=3228)</th>
<th>Qualtrics valid set (n=4389)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Hispanic or Latinx</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>254 (7.9)</td>
<td>0.87 (0.34)</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>7 (0.2)</td>
<td>1.00 (0.00)</td>
</tr>
<tr>
<td>Asian</td>
<td>219 (6.8)</td>
<td>0.92 (0.28)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>728 (22.6)</td>
<td>0.80 (0.40)</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>11 (0.3)</td>
<td>1.00 (0.00)</td>
</tr>
<tr>
<td>White</td>
<td>1889 (58.5)</td>
<td>0.87 (0.34)</td>
</tr>
<tr>
<td>Multiracial or others</td>
<td>120 (3.7)</td>
<td>0.76 (0.43)</td>
</tr>
</tbody>
</table>

**Variations in Survey Responses During Study Period**

The response rate varied throughout the study and was correlated to several social media advertising campaigns and an extended holiday break where no responses were accepted. Figure 3 shows the cumulative number of responses over time, differentiated by fraud detection method and fraud classification. Time periods when social media advertising campaigns were active are highlighted on these plots. Responses tend to increase during social media campaigns. Notably, between January 13 and February 7, 2022, no social media campaign was active, yet a significant number of responses were received (n=766). These responses were largely classified as fraud by our multilayer fraud detection method (n=716, 93.47% classified as fraud) but were often classified as valid by the Qualtrics fraud detection method (n=296, 38.64% classified as fraud). Figure 4 shows the proportion of responses classified as fraud across the study period. The comparative fraud detection between the 2 methods was similar during the first half of the study period (November to December 2021), while the multilayer fraud detection method consistently identified a higher proportion of responses as fraud during the later portion of the study period (January to February 2022).
Figure 3. Fraud detection by multilayer fraud detection system and Qualtrics fraud detection system during the study period (November 2021 to February 2022). Highlighted regions indicate periods when social media campaigns were active (November 3-18, 2021; November 29-December 22, 2021; and February 7-10, 2022. Data collection was paused during an extended winter break from December 23, 2021, until January 12, 2022. The dotted vertical line represents January 12, 2022, when data collection was resumed.

Figure 4. Proportion of responses classified as fraud by multilayer fraud detection system and Qualtrics fraud detection system during the study period (November 2021 to February 2022). Smooth locally weighted smoothing lines are included to help visualize the trends over time.

Discussion

Principal Findings

Our multilayer fraud detection methods identified a substantial number of fraudulent cases. However, when comparing our fraud detection methods to proprietary fraud detection systems provided by web-based survey software, we saw low levels of agreement between the 2 methods. Our results highlight how the choice of fraud detection method can alter the distribution of key study variables.

Both our multilayer fraud detection methods and the Qualtrics fraud detection system identified significant levels of fraud; however, the 2 methods differed in which cases they identified as fraudulent and in which they identified as valid. Notably, these differences were most pronounced for participants who administered the survey specifically for parents. It is possible that fraudulent participants made assumptions about eligibility or compensation (eg, parents are a more specific demographic that may be of interest to researchers, and thus, fraudulent entries that claim to be parents may be more likely to screen as eligible and receive compensation) and responded to eligibility questions in ways that guided them to the parent survey. With a greater
number, and perhaps a greater variety, of fraudulent participants, we then may have seen greater variability in the 2 methods’ ability to consistently identify the fraud.

The disagreement between our multilayer fraud detection and the Qualtrics fraud detection suggests that there are important differences in the methodologies being used by the 2 systems, which resulted in differences in classification. To fully understand and compare the relative performance of the 2 systems, detailed information is needed about the methods used by the Qualtrics fraud detection system. It is likely that the features of the RelevantID proprietary fraud detection method used by Qualtrics are intentionally obfuscated to prevent fraudulent participants from undermining its effectiveness. The trade-off for this black box tactic is that researchers who use the Qualtrics platform cannot ascertain how Qualtrics’ fraud detection algorithms function and how these methods compare to alternative fraud detection strategies. There is an inherent tension between transparency (ie, publishing the features of a fraud detection method improves scientific rigor) and defending against fraud (ie, making comprehensive information about a fraud detection method available may enable fraudulent participants to avoid detection) [22]. Additional research is needed to evaluate the effectiveness of proprietary fraud detection systems and compare them to published techniques used by researchers.

For both fraud detection methods, the cases identified as fraud differed in nonrandom ways from the cases classified as valid on key study variables. These differences could have implications for the interpretation of study results; if legitimate survey participants are classified as fraudulent, critical data are lost, and potential bias could be introduced. In addition, many automated fraud detection tools turn to proprietary machine learning data and predictive modeling for fraud detection [22,38]. This could disproportionately affect those with low literacy or barriers to internet access, as fraud detection tools may be more likely to flag them as fraudulent [19]. Given the existing digital divide across racial or ethnic groups [39,40], this may result in the further exclusion of racial or ethnic minorities in research if they are more likely to use older technologies and shared devices at home or in public spaces, such as computers at libraries or community centers (which may trigger the threshold for fraud due to user behavior that is atypical of a single-user device). One approach to overcoming this challenge is to integrate manual inspection of survey entries in place of or in addition to automatic processes that could result in bias [41,42]. However, human inspection of each survey entry can be time-consuming, more variable than automated processes, and could also result in bias. Future work should aim to better characterize subgroups that may be disproportionally flagged by fraud detection systems and develop ensemble approaches that integrate manual and automatic fraud detection while balancing fraud detection accuracy with protections against excluding valid participants.

While overly sensitive fraud detection could result in bias, fraud detection methods that are not sensitive enough to detect fraudulent entries could also add random noise or systematic bias to the data and threaten the integrity of the research [14-16]. It is important to note that we do not have insight into fraudulent participants’ techniques for responding to survey questions. Fraudulent participants may deliberately select specific demographic options (characteristics they believe will be more likely to result in their entry into the study), randomly select their responses, or use some combination of those techniques [9]. Additionally, rapid developments in machine learning and artificial intelligence have increasingly allowed bots to mimic human behavior [11,22], which could contribute to the seemingly human selection of responses on these surveys, including entries into free text fields [43]. Regardless, this analysis demonstrated the importance of developing study-specific fraud detection methods to supplant or supplement the proprietary fraud detection methods of web-based survey platforms.

Another point of note is the decreased effectiveness of fraud detection tools in determining user legitimacy, as major technology companies take increasing measures to protect user privacy. For example, it is common for fraud detection tools to rely on device fingerprinting and browser cookies to help determine the legitimacy of an individual [44]. While these 2 methods are regularly used by advertisers and marketers to track individuals and deliver targeted advertisements, they also provide a way for fraud detection tools to flag known bad actors and differentiate between legitimate and fraudulent responses. However, the invasive and comprehensive nature of device and browser fingerprinting has raised privacy concerns from users and privacy advocates alike [45,46]. Technology companies, such as Apple, Mozilla, and Brave, have in turn introduced measures to hide users’ identities and activity in a bid to protect user privacy. For instance, Apple’s Safari browser on the macOS desktop operating system now strips all unique identifiers from a user’s device profile, so they appear no different from millions of other Safari users [47]. These privacy-protecting measures, while helpful in safeguarding an individual’s digital presence, make it more difficult for fraud detection tools to differentiate between a legitimate human and a bot. This could partially explain the discrepancy we found between the fraud detection by Qualtrics using reCAPTCHA and RelevantID and our multilayer fraud detection.

Without a method to make a conclusive determination regarding which entries are truly fraudulent and which entries are genuinely valid, it is difficult to compare the relative performance of our multilayer fraud detection methods with the Qualtrics fraud detection methods. However, several pieces of evidence suggest that our fraud detection methods have advantages over Qualtrics in this study context. First, we saw that for email addresses that had an institutional affiliation (ie, “.edu” or “.gov,” which require identity confirmation and cannot be generated en masse) and thus were presumed to be valid, our fraud detection methods correctly validated 98% (n=166) of cases. In comparison, Qualtrics only validated 75% (n=126) of cases. Second, we saw an unusually large discrepancy between the 2 fraud detection methods during a period when the survey link was open, but no advertising or recruitment had recently been active. During this time when we did not expect to receive legitimate responses, we received hundreds of responses that were largely classified as fraud by our fraud detection methods but were generally classified as valid by the Qualtrics system.
While it is possible that valid participants were still able to find and access this survey in the absence of active recruiting, we believe this pattern is evidence of noneligible actors using automated systems in an attempt to gain additional compensation payments from the survey. Taken together, these 2 observations are indirect evidence that our multilayer fraud detection method may have better specificity (ie, can correctly identify valid entries) as well as better sensitivity (ie, can accurately detect fraudulent entries) when compared to the Qualtrics system in this study. While we are unable to conclude whether the approach we developed for this study is more or less accurate in identifying fraud when compared to the system used by Qualtrics, we believe these pieces of indirect evidence suggest that using an automated system, such as the one available through Qualtrics, alone may be suboptimal. A multilayered approach was recommended to fraud detection that includes conservative use of automated fraud detection and integration of human review of entries that is tailored to the study’s specific context and its participants.

Limitations
This study is subject to several limitations. First, our comparison of fraud detection methods is limited by the fact that we are unable to definitively determine which entries are valid and which are fraudulent. We selected fraud detection criteria specifically intended to identify repeat respondents (eg, multiple responses providing identical information), fraudulent submissions from outside the Philadelphia region (eg, location verification using IP addresses), and submissions from bots or bot-assisted fraudulent participants (eg, requiring responses that would be difficult to generate via algorithm such as local, colloquial neighborhood names). Applying these criteria may still have resulted in the inclusion of illegitimate responses and the exclusion of legitimate ones. Second, because we cannot know for certain the true fraud status of participants, we are unable to calculate metrics like precision and recall for the fraud detection methods. Future research should aim to establish gold-standard indicators for fraud that could then be used to directly compare the efficacy of the different methods for fraud detection. Third, without knowing how Qualtrics detects fraud, we are unable to determine which components of our strategy may overlap with the Qualtrics strategy. This limits the conclusions we can draw about the comparative effectiveness of these fraud detection methods. Fourth, while we choose to compare our fraud detection methods with the automated systems used by the Qualtrics platform, we believe that similar comparisons and research are also needed with other proprietary fraud detection systems.

Recommendations
The following recommendations are offered for improving data integrity in web-based survey research based on the findings from this study:

- Use a multilayered approach to fraud detection that combines different techniques like bot detection, location verification, consistency checks, and manual review. Relying solely on one method may miss certain types of fraud.
- Carefully evaluate proprietary fraud detection systems and request details on their methodology if possible. Black box methods make it difficult to fully assess their impact on sample composition.
- Avoid overly strict fraud detection rules that may disproportionately exclude valid respondents from vulnerable groups. Balance rigor with inclusion.
- Continuously monitor survey responses over time to identify changes in fraud patterns that may require adjustments to detection methods.

Conclusions
Web-based research and recruitment through social media platforms offer powerful flexibility for researchers to collect large, diverse samples. Web-based surveys, however, are vulnerable to low-quality data from fraud and duplicate entries. Researchers must actively design their web-based studies with this vulnerability in mind and adopt active and adaptable methods of detecting and responding to fraudulent survey responses. Automated, proprietary fraud detection systems offered by web-based survey software may be an important tool in combating fraud, but additional research is needed to evaluate their effectiveness. Human verification of survey entries, while time-consuming, can add another layer of protection and enhance the rigor of web-based survey research. We believe a multilayered strategy that includes a combination of automated fraud detection tools, data enrichment, and human intelligence is the best approach for combating fraud.

Acknowledgments
The authors would like to recognize the contributions of the community partners who helped to distribute the survey as well as the research participants who shared their experiences and insights as part of this research. The authors would also like to thank Gared Harbison for his contributions to validating survey responses and cleaning and managing the data. This work was supported by funding from the National Institutes of Health Agreement OT2HL16156 as part of the Community Engagement Alliance Against COVID-19 Disparities, PhillyCEAL.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.
Authors' Contributions

SB, PST, JW, AD, AV, and JB conceptualized the research project; SB, PST, JW, and AD cleaned and managed the data; SB conducted formal statistical analysis of the data; AV and JB acquired financial support for the project; SB, PST, JW, JG, and JB developed the methodology for the analysis; SB, PST, and JW managed and coordinated day to day operations for the project; AV and JB provided oversight and leadership for the research; SB and WL created visualizations and data presentations; and SB, WL, PST, JW, and JG wrote the original draft of the paper. All authors reviewed and revised the paper and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Checklist for Reporting Results of Internet E-Surveys (CHERRIES).

[DOCX File . 22 KB - formative_v811e47091_app1.docx ]

Multimedia Appendix 2

Fraud detection analysis.

[DOCX File . 39 KB - formative_v811e47091_app2.docx ]

References

27. reCAPTCHA v2. Google. 2023. URL: https://developers.google.com/recaptcha/docs/display [accessed 2023-12-14]
33. Proxy and VPN detection API. proxycheck. 2022. URL: https://www.proxycheck.io [accessed 2023-12-14]


©Stephen Bonett, Willey Lin, Patrina Sexton Topper, James Wolfe, Jesse Golinkoff, Aayushi Deshpande, Antonia Villarruel, José Bauermeister. Originally published in JMIR Formative Research (https://formative.jmir.org), 12.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

https://formative.jmir.org/2024/1/e47091

JMIR Form Res 2024 | vol. 8 | e47091 | p.480

Paywall-Access

https://formative.jmir.org/2024/1/e47091

(abstract not for citation purposes)
Virtual and Interprofessional Objective Structured Clinical Examination in Dentistry and Dental Technology: Development and User Evaluations

MengWei Pang¹²³⁴*, MDS; YanLing Dong¹²³⁴*, MDS; XiaoHan Zhao⁵*, MD; JiaWu Wan⁶, BA; Li Jiang¹²³⁴, DT; JinLin Song¹²³⁴, PhD; Ping Ji¹²³⁴, PhD; Lin Jiang¹²³⁴*, DDS

¹Stomatological Hospital of Chongqing Medical University, Chongqing, China
²Chongqing Key Laboratory of Oral Diseases and Biomedical Sciences, Chongqing, China
³Chongqing Municipal Key Laboratory of Oral Biomedical Engineering of Higher Education, Chongqing, China
⁴College of Stomatology, Chongqing Medical University, Chongqing, China
⁵State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing, China
⁶Beijing Unidraw Virtual Reality Technology Research Institute Co Ltd, Beijing, China
*these authors contributed equally

Corresponding Author:
Lin Jiang, DDS
Stomatological Hospital of Chongqing Medical University
426# Songshibei Road
Yubei District
Chongqing, 401147
China
Phone: 86 15922650133
Email: jianglin@hospital.cqmu.edu.cn

Abstract

Background: Interprofessional education (IPE) facilitates interprofessional collaborative practice (IPCP) to encourage teamwork among dental care professionals and is increasingly becoming a part of training programs for dental and dental technology students. However, the focus of previous IPE and IPCP studies has largely been on subjective student and instructor perceptions without including objective assessments of collaborative practice as an outcome measure.

Objective: The purposes of this study were to develop the framework for a novel virtual and interprofessional objective structured clinical examination (viOSCE) applicable to dental and dental technology students, to assess the effectiveness of the framework as a tool for measuring the outcomes of IPE, and to promote IPCP among dental and dental technology students.

Methods: The framework of the proposed novel viOSCE was developed using the modified Delphi method and then piloted. The lead researcher and a group of experts determined the content and scoring system. Subjective data were collected using the Readiness for Interprofessional Learning Scale and a self-made scale, and objective data were collected using examiner ratings. Data were analyzed using nonparametric tests.

Results: We successfully developed a viOSCE framework applicable to dental and dental technology students. Of 50 students, 32 (64%) participated in the pilot study and completed the questionnaires. On the basis of the Readiness for Interprofessional Learning Scale, the subjective evaluation indicated that teamwork skills were improved, and the only statistically significant difference in participant motivation between the 2 professional groups was in the mutual evaluation scale (P=.004). For the viOSCE evaluation scale, the difference between the professional groups in removable prosthodontics was statistically significant, and a trend for negative correlation between subjective and objective scores was noted, but it was not statistically significant.

Conclusions: The results confirm that viOSCE can be used as an objective evaluation tool to assess the outcomes of IPE and IPCP. This study also revealed an interesting relationship between mutual evaluation and IPCP results, further demonstrating that the IPE and IPCP results urgently need to be supplemented with objective evaluation tools. Therefore, the implementation of viOSCE as part of a large and more complete objective structured clinical examination to test the ability of students to meet undergraduate graduation requirements will be the focus of our future studies.
KEYWORDS
dentist; dental technician; objective structured clinical examination; OSCE; interprofessional education; interprofessional collaborative practice

Introduction

Interprofessional Collaboration Between Dentists and Dental Technicians

Conflicts are part of the life of any organization, and the dental professions are not spared. Jurisdictional battles and supremacy struggles are not alien to dentistry [1]. Unfortunately, despite the obvious reported benefits of interprofessional education (IPE) for interprofessional collaborative practice (IPCP) [2], there is a paucity of data about IPE to promote IPCP among dental professionals. Dentists and dental technicians need to communicate effectively and contribute their professional skills to ensure that they make decisions that are in the best interests of their patients [3]. A clear understanding of the interactions of the dental team can promote teamwork [4], establish cooperative goals [5], encourage mutual respect [6], and promote IPCP between dental students and dental technology students [7].

IPE encourages teamwork among dental care professionals [8-11] and is increasingly becoming a part of the training programs for dental and dental technology students [12-17]; however, gaps still remain. Perhaps, the largest gaps are owing to the predominant focus of previous studies regarding IPE and IPCP on student and instructor perceptions and a lack of objective assessment of collaborative practice as an outcome measure [18,19]. This marked gap has necessitated the development of a conceptual framework to evaluate the impact of IPE on IPCP to strengthen the evidence for IPE as a tool to improve IPCP between dental and dental technology students [20].

The Objective Structured Clinical Examination

The objective structured clinical examination (OSCE) is an assessment tool based on the principles of objectivity and standardization, in which individual students move through a series of time-limited stations in a circuit for the purpose of assessment of professional performance in a simulated environment. At each station, the student is assessed and marked against standardized scoring rubrics by trained assessors [21]. OSCE has been widely adopted as a summative assessment in the medical undergraduate curriculum and is universally accepted as the gold standard for assessing clinical competence in dental education [22,23]; furthermore, its effectiveness has been confirmed by several studies [24-26]. On the basis of the extensive application of OSCE, the interprofessional OSCE (iOSCE) was initially developed to simulate IPCP [27]. Unlike conventional OSCE, iOSCE involves students from different professions, encourages students to work as a team, and requires the entire team to participate in all tasks [28]. This is performed to objectively evaluate the results of IPE. Within this framework, several variations of iOSCE have been developed to accommodate the training needs of health care teams built to address different disease categories (team OSCE [29,30], group OSCE [31,32], interprofessional team OSCE [28,33], etc). These iOSCE variants can be roughly divided into synchronous [33-36] and asynchronous [29,37] task-based variants. A team working in an operating room typically works synchronously, whereas health care teams of dentists and dental technicians typically work asynchronously. Although the use of iOSCE in medical education has been extensively reported [27-30,38-42], to the best of our knowledge, the use of iOSCE for asynchronous work, especially within dentistry and dental technology cross-professional education, has not been reported.

Although iOSCE may provide an ideal solution for dental and dental technology students to perform IPCP simulation based on real patient cases, the COVID-19 pandemic [43] highlighted the limitations of this traditional approach. For example, a plaster model generated from a clinical case and passed multiple times among students and examiners may pose a risk of infection. In addition, diagnostic stations are usually set up to facilitate OSCE. A station is typically equipped with a trained, standardized patient, and the students complete the diagnosis by asking questions and examining this standardized patient. The risk of infection at this type of station was heightened during the pandemic. Nevertheless, compared with the traditional OSCE, iOSCEs are more time consuming and resource intensive [44,45], especially in dental education; hence, a virtual approach, as developed and piloted in this study, is justified [46,47]. Notably, the conventional virtual OSCE (vOSCE) has been described as a method of performing OSCE using internet technology in medicine [47-49]. The major reason for this technological approach was the scattered nature of the locations of students requiring assessment. However, this approach does not fully leverage virtual technology in dentistry. The integration of digital dental technologies and cloud-based dental laboratory workflows could be practiced within the vOSCE framework [50], which now also forms a professional core course in dental technology education [51-53]. The development of iOSCE based on virtual technology could facilitate the inclusion of digital dental technology in the blueprint design of examination stations. This combination could simulate the critical needs of present-day dental laboratories and promote students’ improved perception about the current demands of the profession.

Objective

To address these research gaps, this study presented a new virtual iOSCE (viOSCE) to objectively assess the effectiveness of IPE as a tool to promote IPCP among dental and dental technology students. We have described the development and piloting of a viOSCE framework and its virtual techniques to validate the user-friendliness of IPE and document its effect on IPCP among dental and dental technology students. Data from both subjective and objective evaluations were collected, and their correlation was assessed.

https://formative.jmir.org/2024/1/e44653
Methods

Development of viOSCE

The principal investigator (PI) first limited the viOSCE knowledge to content related to the prosthodontics course. Content related to implantology and orthodontics was excluded because it is not part of the core undergraduate coursework for dental or dental technology students. On the basis of the Association for Medical Education in Europe guide [54], a modified Delphi method was used to generate content for viOSCE. The Delphi method is a decision-making process that uses expert opinion, gathered in the form of a survey, under the guidance and direction of the PI to reach group consensus through collaboration, independent analysis, and iteration [55]; this process is the most frequently used method to generate content for OSCEs [54]. The panel of experts in this study consisted of 9 instructors (including the PI) from the College of Stomatology, Chongqing Medical University. All 9 instructors had prosthodontics teaching experience and digital technology practical teaching experience with undergraduate dental and dental technology students. They had also participated in the design and examiner training for traditional OSCE, but only the PI had experience in IPE and vOSCE design.

In this study, there were 4 iterations (rounds) before the viOSCE station design was finalized. In the first round, the PI identified 10 potential topics for viOSCE based on the syllabus of the prosthodontics course for dentistry and dental technology students, gave initial suggestions for the station design, and created a manuscript that was emailed to the panel of experts. Each expert independently gave their opinion and selected 5 topics that they considered as the most important in the syllabus and the most suitable for assessment using viOSCE. In the second round, the PI identified 3 topics with the highest selection rate based on the expert feedback and designed draft blueprints for 20 stations based on the top 3 selected topics using existing virtual technology support. These were sent to the expert panel via email. The expert panel commented about the potential effectiveness of interprofessional collaboration at the stations, made necessary corrections, and returned the design drafts to the PI. In the third round, the PI summarized all the changes made by the expert panel and, finally, decided on 7 stations based on the availability of virtual technology and the time to be spent on the stations within the allotted time frame of the examination. Stations consuming a lot of time, requiring multiple devices for support, or requiring very large spaces were rejected. Next, the selected viOSCE station blueprint design was completed, the virtual technical support was finalized, and the PI sent the final viOSCE station blueprint to the expert panel via email. The expert panel created the scoring rubrics based on the final viOSCE station blueprint, and these were returned to the PI for finalization. In the final round, the PI compiled all the information and met with the group to get a consensus regarding the viOSCE station blueprint and scoring rubrics. Once all the experts approved the viOSCE test station blueprint and scoring rubrics, the PI declared the viOSCE design as complete and declared the panel of experts the viOSCE examiner panel (Figure 1).
The viOSCE development process based on the modified Delphi method. PI: principal investigator; viOSCE: virtual and interprofessional objective structured clinical examination.

**Figure 1.** The viOSCE development process based on the modified Delphi method. PI: principal investigator; viOSCE: virtual and interprofessional objective structured clinical examination.

The viOSCE Framework

The developed viOSCE framework consisted of 3 topics, namely, fixed prosthodontics, removable prosthodontics, and clinical diagnostics. There were 7 collaborative examination stations consisting of 4 asynchronous and 3 synchronous stations. All these stations were designed and developed using the Delphi method (Figure 2).
At the fixed prosthodontic stations, the dental student prepared tooth 8 (maxillary right central incisor) on the simulator (Nissan Dental Products) and then worked with the dental technology student to scan the preparations using an intraoral scanner (Panda P2; Freqty Technology). The dental and dental technology students at the intraoral scanning station worked collaboratively. The dental student performed an intraoral scan task, and the dental technology student observed the scan results to determine whether they could be used for the computer-aided design (CAD) wax pattern station. After obtaining a digital model, the dental technology student used a CAD system (Dental system; 3shape) to design a single crown on the digital model of the preparation. Individual scoring rubrics were designed for tooth preparation, intraoral scan, and CAD wax pattern. The 3 examiners scored each of the 3 stations (Figure 3 and Tables 1-3).
Figure 3. The fixed prosthodontics stations of the viOSCE. (A) A dental student prepared tooth 8 on the simulator. (B) Dental and dental technology students scanned the preparation using an intraoral scanner. (C) A dental technology student created a digital wax pattern using the computer-aided design system. (D) A viOSCE examiner scored the preparation process and results. (E) A viOSCE examiner scored the intraoral scanning process and results. (F) A viOSCE examiner scored the digital wax patterns. viOSCE: virtual and interprofessional objective structured clinical examination.
Table 1. The scoring rubric used to assess the tooth preparation stations.

<table>
<thead>
<tr>
<th>Scoring component</th>
<th>Points</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preparation before operation</td>
<td>5</td>
<td>• Infection control was correctly performed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct adjustment of phantom head position and lighting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The operating position is correct</td>
</tr>
<tr>
<td>Fine motor skills</td>
<td>15</td>
<td>• Holds the handpiece correctly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Fulcrum stability</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct use of the mouth mirror to reflect areas to be operated under indirect vision</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Accurate application of burs</td>
</tr>
<tr>
<td>Preparation during operation</td>
<td>15</td>
<td>• Operation sequence correctly performed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Placement of depth orientation grooves</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Labial surface prepared in 2 planes</td>
</tr>
<tr>
<td>Incisal reduction</td>
<td>10</td>
<td>• 1.5-2 mm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Formed a small bevel inclined 45° to the lingual side</td>
</tr>
<tr>
<td>Axial reduction</td>
<td>15</td>
<td>• 2 mm for the labial surface</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 1 mm for the proximal surface</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 0.7-1 mm for the lingual surface</td>
</tr>
<tr>
<td>2-plane reduction</td>
<td>5</td>
<td>• Labial surface forms 2 planes and has rounded line angles and point angles</td>
</tr>
<tr>
<td>Taper</td>
<td>5</td>
<td>• Retentive walls: 6°-10°</td>
</tr>
<tr>
<td>Margin placement</td>
<td>10</td>
<td>• Margins extended to a specified target (1 mm supragingivally)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 0.8-1 mm for the shoulder, modified form of the shoulder, and small radius internal angle with a 90° cavosurface margin</td>
</tr>
<tr>
<td>Details</td>
<td>20</td>
<td>• Adjacent teeth and gingiva are unaffected by the preparation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No undercut areas</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Margins and walls are smooth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Margins are continuous and well defined</td>
</tr>
</tbody>
</table>

Table 2. The scoring rubric used to assess the intraoral scan stations.

<table>
<thead>
<tr>
<th>Scoring component</th>
<th>Points</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scanning preparation</td>
<td>25</td>
<td>• Order creation is correct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The tip is held smoothly and stable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No saliva interference during scanning</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Cleans the lens and waits for 10 s to preheat the lens</td>
</tr>
<tr>
<td>Scanning operation</td>
<td>35</td>
<td>• Continuous operation of the standard scanning sequence without pauses</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• During the scanning, the lip and other soft tissues are pulled to expand the scanning field</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Scanning should be completed in 6 min (upper and lower jaws)</td>
</tr>
<tr>
<td>Scanning integrity</td>
<td>35</td>
<td>• Mesial and distal interproximal surfaces are intact with no missing red-blue data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The scan width of the gingival area is at least 2 mm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Scanning of the occlusal surface or incisal edge is complete and clear</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The bite registration is correct</td>
</tr>
<tr>
<td>Software tool selection</td>
<td>5</td>
<td>• Ability to use the software tools accurately</td>
</tr>
</tbody>
</table>
Table 3. The scoring rubric was used to assess the computer-aided design wax pattern stations (crowns).

<table>
<thead>
<tr>
<th>Scoring component</th>
<th>Points</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order creation</td>
<td>5</td>
<td>• Selects preparation in the teeth overview correctly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Selects the category correctly (anatomy, wax, and zirconia)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Selects the import option correctly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Selects the relevant import type correctly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Imports the intraoral scan data correctly</td>
</tr>
<tr>
<td>Margin</td>
<td>10</td>
<td>• Places the margin line correctly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Sets the insertion direction correctly</td>
</tr>
<tr>
<td>Occlusion</td>
<td>15</td>
<td>• Normal overlap and overbite</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Accurate restoration of the occlusal vertical dimension</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The occlusion can be checked by dynamic virtual articulation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Balanced occlusal forces and no premature contacts</td>
</tr>
<tr>
<td>Proximal contact area</td>
<td>10</td>
<td>• Correct position and shape of the proximal contact area</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct contact relationship between adjacent teeth</td>
</tr>
<tr>
<td>Shape</td>
<td>35</td>
<td>• Tooth position: long axis is correctly aligned with the lip and tongue</td>
</tr>
<tr>
<td></td>
<td></td>
<td>direction, correct proximal and distal orientation, tooth is correctly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>positioned in the dental arch, and ratio of the tooth length to width</td>
</tr>
<tr>
<td></td>
<td></td>
<td>is coordinated with that of the adjacent teeth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Thickness: the thinnest thickness is not &lt;0.5 mm, and the axial surface</td>
</tr>
<tr>
<td></td>
<td></td>
<td>thickness is not &lt;1 mm and not &gt;1.5 mm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Gingival embrasures are correctly designed and coordinated with</td>
</tr>
<tr>
<td></td>
<td></td>
<td>those of the adjacent teeth</td>
</tr>
</tbody>
</table>
|                        |        | • Tooth length: the incisal position is in harmony with that of the adja-
|                        |        | cent teeth                                                              |
|                        |        | • Detailed structure of the surface, such as developmental grooves and  |
|                        |        | ridges                                                                  |
|                        |        | • Gingival embrasures are correctly designed and coordinated with       |
|                        |        | those of the adjacent teeth                                             |
|                        |        | • Lingual morphology: lingual fossa and marginal ridge morphology       |
| Cement space           | 5      | • Acceptable cement space                                               |
|                        |        | • Acceptable extra cement space                                        |
| Restoration effect     | 20     | • Acceptable functionality                                              |
|                        |        | • Acceptable esthetics                                                  |
|                        |        | • Acceptable visual harmony                                             |

At the removable prosthodontics station, real patient cases and intraoral digital models were selected and prepared by the PI, followed by approval by the expert panel. The intraoral digital model was a clinical plaster model scanned using Lab Scanner (E4; 3shape). Each dental student used our previously developed Objective Manipulative Skill Examination of Dental Technicians (OMEDT) system [56] to observe the intraoral digital model and to design a removable partial denture (RPD) framework. At the end of the design task, the dental student submitted the design and then discussed the design with the dental technology student; the dental student could make modifications if they wanted to. Next, each dental technology student used a CAD system (Dental system; 3shape) to design the framework of an RPD on the intraoral digital model based on the final design. A viOSCE examiner scored the first RPD design using the OMEDT system. Next, the viOSCE examiner scored the final RPD design and the digital framework of the RPD. The design discussion station was not scored by a separate examiner (Figure 4 and Tables 4 and 5).
Figure 4. The removable prosthodontics stations of the viOSCE. (A) A dental student designed the framework of a RPD using the Objective Manipulative Skill Examination of Dental Technicians system. (B) Dental and dental technology students discussed the RPD design. (C) A dental technology student created a digital framework of an RPD using the computer-aided design system. (D) A viOSCE examiner scored the first RPD design using the Objective Manipulative Skill Examination of Dental Technician system. (E) A viOSCE examiner scored the final RPD design and the digital framework of the RPD. RPD: removable partial denture; viOSCE: virtual and interprofessional objective structured clinical examination.

Table 4. The scoring rubric used to assess the removable partial denture design stations.

<table>
<thead>
<tr>
<th>Scoring component</th>
<th>Points</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case observation</td>
<td>20</td>
<td>The missing tooth position is identified accurately and marked correctly on the drawing</td>
</tr>
<tr>
<td>Design choices</td>
<td>40</td>
<td>No missing component</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Indirect retainer is present in the optimal position</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Design choices do not violate biological principles</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clasp choice is optimal for the case</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The major connector is selected properly with reasonable extension</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Justified use of clasps and rests</td>
</tr>
<tr>
<td>Drawing</td>
<td>20</td>
<td>Ideal drawing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metal components are painted in blue, and resin bases are painted in red</td>
</tr>
<tr>
<td>Consistency with task description</td>
<td>10</td>
<td>Exactly as described in the task description</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clearly presents the requirements implied in the description, and the design is well aligned with the corresponding description</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gives consideration to both esthetics and functions</td>
</tr>
<tr>
<td>Neatness and accuracy in presentation</td>
<td>10</td>
<td>Neat and accurate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No inconsistencies between the table and drawing</td>
</tr>
</tbody>
</table>
Table 5. The scoring rubric used to assess the computer-aided design wax pattern stations (removable partial denture [RPD] framework).

<table>
<thead>
<tr>
<th>Scoring component</th>
<th>Points</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order creation</td>
<td>5</td>
<td>• Selects artificial teeth in the teeth overview correctly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Selects the category correctly (removable—RPD frame)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Selects the import option correctly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Selects the relevant import type correctly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Imports the laboratory scan data correctly</td>
</tr>
<tr>
<td>Surveying</td>
<td>10</td>
<td>• Insertion direction is correctly chosen</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Undercuts are correctly identified</td>
</tr>
<tr>
<td>Virtual cast preparation</td>
<td>20</td>
<td>• Correct paralleled blockout</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct shaped blockout</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct arbitrary blockout</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct relief setting</td>
</tr>
<tr>
<td>Framework design</td>
<td>40</td>
<td>• Reasonable position and shape of clasp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Reasonable position and shape of rest</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Reasonable position and shape of major connector</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Reasonable position and shape of retention grid</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Reasonable position and shape of finishing line</td>
</tr>
<tr>
<td>Form</td>
<td>25</td>
<td>• All parts are connected as a whole</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The thickness and strength of the framework meet the requirements</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The thickness is uniform, and the surface is smooth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Esthetics are acceptable</td>
</tr>
</tbody>
</table>

The clinical diagnostics station used a virtual standardized patient (VSP) with the haptic device (UniDental, Unidraw). The VSP hardware does not have an anthropomorphic shape, but it interacts through vocal, visual, and haptic devices. On the basis of the novel oral knowledge graph and the coupled, pretrained Bert models, the VSP can accurately interact with a dentist’s underlying intention and express the symptom characteristics in a natural style [57]. On the basis of this algorithm, the PI adjusted and entered the real patient case details, allowing the dental technology student to work with the dental student as a chairside dental technician to make a diagnosis based on the information obtained from the interactions with the VSP. In this study, the clinical case designed on the VSP was a patient who required root canal treatment and full crown restoration. At the end of the dental student’s diagnosis and simulation, the dental technology student was required to assist the dental student in designing the restoration plan and help the patient in choosing the materials for crown restoration (this often determines the price of the treatment). Thus, dental and dental technology students finalized the prosthodontic treatment plan collaboratively. The visual device built a virtual dental clinic environment and VSP model, allowing the students to view the VSP from global, extraoral, and intraoral perspectives. The haptic device allows dental students to perform intraoral and extraoral examinations using essential tools to explore the diagnostic evidence.

Owing to the complexity of collaborative diagnosis, the station was manually scored by 2 examiners independently based on the previously developed scoring rubrics, whereas the UniDental output machine provided an additional score according to the previously developed scoring rubrics. The average of the 3 scores formed the final score for the station. To ensure the relative independence and internal consistency of all scores, the examiners were not informed about the existence of the machine score. The PI exported the machine score data from the VSP at the end of the experiment (Figure 5 and Table 6).
Figure 5. The clinical diagnostics station of the viOSCE. (A) The VSP with the haptic device, UniDental. (B) Dental and dental technology students performed intraoral palpation on the VSP using the haptic device. (C) Then, 2 viOSCE examiners scored the process and clinical diagnostic results. viOSCE: virtual and interprofessional objective structured clinical examination; VSP: virtual standardized patient.
Table 6. The scoring rubric used to assess the clinical diagnostic stations.

<table>
<thead>
<tr>
<th>Scoring component</th>
<th>Points</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>History taking</td>
<td>25</td>
<td>• The content of the inquiry is accurate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Few questions unrelated to the disease or clinical situation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Inquiries are made sequentially, purposefully, and hierarchically</td>
</tr>
<tr>
<td>Intraoral examination</td>
<td>25</td>
<td>• Tool selection is accurate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Humanistic care is reflected during the examination</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Appropriate oral examination items are performed based on the case information</td>
</tr>
<tr>
<td>Auxiliary examination</td>
<td>10</td>
<td>• Correct auxiliary examination items are selected</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct interpretation of auxiliary examination results</td>
</tr>
<tr>
<td>Case analysis</td>
<td>20</td>
<td>• Correct diagnosis of the case</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct selection of the diagnostic criteria</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct differential diagnosis of the case</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct selection of the basis for differential diagnosis</td>
</tr>
<tr>
<td>Plan design</td>
<td>20</td>
<td>• Correct treatment plan design according to the disease condition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Provides advice about material selection based on patient request</td>
</tr>
</tbody>
</table>

Performance Evaluation of viOSCE

In this study, fourth-year undergraduate dental students and third-year undergraduate dental technology students participated in viOSCE because students at this stage of education had completed preclinical professional training. Overall, 50 students who met these requirements, including 25 (50%) dental students and 25 (50%) dental technology students, were recruited into the viOSCE user evaluation experiment and were divided into groups of 2 comprising 1 dental student and 1 dental technology student. The PI and examiner teams did not influence or determine the team-formation process. All participating students were informed that as this viOSCE was in the experimental phase, it was conducted as a small extracurricular skills competition, thus allowing for self-evaluation without a final examination situation, as previously reported [58]. This approach allowed for the simulation of an examination situation without affecting the final examination grade of the students. A month before commencing the experiment, the PI led an web meeting for students to explain the viOSCE, the relevant knowledge points, and the need to practice fully during the upcoming month. At the end of the meeting, the students completed the Readiness for Interprofessional Learning Scale (RIPLS) pretest questionnaire, which is a 19-item 5-point Likert-scale questionnaire; this type of questionnaire is the most frequently used method for the subjective evaluation of IPE and IPCP [18].

viOSCE was piloted after the 1-month preparation period. The panel of examiners marked points according to the previously prepared scoring rubrics, whereas some of the points were automatically scored by a machine. After this step, the participating students completed the posttest self-made questionnaire, to which a mutual evaluation scale and a viOSCE evaluation scale were added. The mutual evaluation scale asked the students to score the performance of their partner, whereas the viOSCE evaluation scale asked the students to score viOSCE. In total, 6 items were included in the mutual evaluation scale, and 7 items were included in the viOSCE evaluation scale (Textboxes 1 and 2). All items in both questionnaires were set to a maximum score of 100. Before issuing the questionnaire, the panel first reviewed all the questions, clarified ambiguities, and removed any double-barreled questions [59,60]. At the end of the experiment, one-on-one interviews were conducted with all the students to determine their perceptions about viOSCE.

Textbox 1. The mutual evaluation scale administered to dental and dental technology student groups who participated in the virtual and interprofessional objective structured clinical examination (viOSCE).

<table>
<thead>
<tr>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final contribution</td>
</tr>
<tr>
<td>Person-organization fit</td>
</tr>
<tr>
<td>Performance in viOSCE</td>
</tr>
<tr>
<td>Professional skill</td>
</tr>
<tr>
<td>Practice volume before viOSCE</td>
</tr>
<tr>
<td>Motivation to participate</td>
</tr>
</tbody>
</table>
The virtual and interprofessional objective structured clinical examination (viOSCE) evaluation scale administered to dental and dental technology student groups who participated in viOSCE.

<table>
<thead>
<tr>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Evaluation of viOSCE effectiveness</td>
</tr>
<tr>
<td>• Evaluation of equipment, network operation and maintenance</td>
</tr>
<tr>
<td>• Evaluation of viOSCE examiners</td>
</tr>
<tr>
<td>• Evaluation of viOSCE staff</td>
</tr>
<tr>
<td>• Rationality of the clinical diagnostic design</td>
</tr>
<tr>
<td>• Rationality of the fixed prosthodontics design</td>
</tr>
<tr>
<td>• Rationality of the removable prosthodontics design</td>
</tr>
</tbody>
</table>

Statistical Analysis

Data were tabulated in a Microsoft Excel spreadsheet and imported into IBM SPSS Statistics for Windows (version 26.0; IBM Corp) for descriptive analysis. GraphPad PRISM 8.0 software (GraphPad Software) was used to create the graphs. Responses were summarized, and comparisons were made. Output data were presented as percentages and in graphical format. The Shapiro-Wilk test was used to test for normal distribution. Specific data analysis tests performed included descriptive statistics, 2-tailed paired t tests, and correlation analyses.

Ethical Considerations

The research ethics committee of the Affiliated Hospital of Stomatology, Chongqing Medical University, approved this study protocol (COHS-REC-2022; LS number: 096). All participants provided written informed consent before participation in the study.

Results

Of the 50 students, 32 (64%) completed the experiment. Interviews were conducted with the students who dropped out of further participation in the study. The main reasons for dropping out included the students’ belief that they or their collaborating partners had not practiced sufficiently to perform well in the experiment. A group had a verbal confrontation approximately an hour before the experiment began. The main reason for the conflict was that the dental technology student accused the dental student of not practicing sufficiently before the experiment. According to the study protocol, at the end of the experiment, the conflict was resolved by the PI. Both parties were counseled, mediated by the PI, and the 2 parties reconciled.

Data from the RIPLS, mutual evaluation scale, and viOSCE evaluation scale were first analyzed to determine the impact of viOSCE on the subjective evaluation of IPCP. All students (32/32, 100%) who completed the experiment were administered the RIPLS questionnaire before and after the experiment. The Cronbach $\alpha$ values were .835 for the pretest data and .731 for the posttest data, suggesting that the reliability and internal consistency were acceptable. The results failed the Shapiro-Wilk test for normality; therefore, the data were analyzed using the Wilcoxon signed rank test. The teamwork and collaboration subscale scores were significantly increased after the experiment ($P=.004$). In addition, there was a nonsignificant decrease in the negative professional identity subscale scores ($P=.21$). There was also an insignificant increase in the scores on the positive identity subscale and on the roles and responsibilities subscale ($P=.13$ and $P=.96$, respectively). Figure 6 depicts the RIPLS data before and after the viOSCE pilot.
After the experiment, the mutual evaluation scale was administered to all participating students (32/32, 100%) who completed the experiment. The Cronbach $\alpha$ value was .873, suggesting good reliability and internal consistency. Comparison of the results of the dental and dental technology students revealed that only the mutual evaluation scores for competition motivation were significantly different between the 2 groups ($P=.04$). The dentistry and dental technology students evaluated each other’s motivation to participate in the competition (competition motivation), and the dental students had higher scores than the dental technology students. Figure 7 depicts the mutual evaluation scale scores of the dental and dental technology students.
Similarly, after the experiment, the viOSCE evaluation scale was administered to all students (32/32, 100%). The Cronbach α value was .706, suggesting acceptable reliability and internal consistency. Comparison of the viOSCE evaluation scale results of the dental and dental technology students with the Wilcoxon signed rank test results revealed that only the evaluation scores for the removable prosthodontics design were statistically significant (P=.01) among the 7 items. Figure 8 depicts the viOSCE evaluation scale scores of the dental and dental technology students.

**Figure 7.** Mutual evaluation scale data for the dental and dental technology students. viOSCE: virtual and interprofessional objective structured clinical examination.

**Figure 8.** viOSCE evaluation scale data for dental and dental technology students. viOSCE: virtual and interprofessional objective structured clinical examination.
To explore the validity of the examiner panel scores in viOSCE, correlation analysis was conducted on the scores of each station under the 3 topics. Using Spearman correlation coefficient, for the fixed prosthodontics topic, a strong positive correlation between the scores of the tooth preparation station and the CAD wax pattern station was noted, and it was statistically significant \((r=0.67; P=.005)\). Positive correlations between the scores of the intraoral scan station and the CAD wax pattern station and between the intraoral scan station and the tooth preparation station were not statistically significant \((r=0.179; P=.51\) and \(r=0.387; P=.14\), respectively). For the removable prosthodontics topic, 11 (69%) of the 16 student groups finally decided to modify the RPD design initially made by the dental students. A negative but statistically insignificant correlation between the scores of the RPD design station and the CAD wax pattern station was noted \((r=-0.111; P=.68)\). For the clinical diagnostics topic, the correlation analysis was conducted primarily for the machine scores and the examiner scores to determine the usability of the VSP in viOSCE and the consistency of machine scoring and examiner scoring. The results revealed a significant positive correlation between the scores of the 2 examiners, and the positive correlation between the machine scores and the 2 examiners’ scores was also significant. The results are shown in Table 7 and Figure 9.

### Table 7. Spearman correlation analysis of the virtual and interprofessional objective structured clinical examination scores.

<table>
<thead>
<tr>
<th>Topic and station</th>
<th>Correlation coefficient</th>
<th>(P) value (2-tailed)</th>
<th>Participants (n=16), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed prosthodontics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tooth preparation vs CAD(^a) wax pattern</td>
<td>0.670</td>
<td>.005(^b)</td>
<td>16 (100)</td>
</tr>
<tr>
<td>Intraoral scan vs tooth preparation</td>
<td>0.387</td>
<td>.14</td>
<td>16 (100)</td>
</tr>
<tr>
<td>Intraoral scan vs CAD wax pattern</td>
<td>0.179</td>
<td>.51</td>
<td>16 (100)</td>
</tr>
<tr>
<td><strong>Removable prosthodontics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPD(^b) design vs CAD wax pattern</td>
<td>–0.111</td>
<td>.68</td>
<td>16 (100)</td>
</tr>
<tr>
<td><strong>Clinical diagnostics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Machine score vs examiner-1 score</td>
<td>0.601</td>
<td>.01(^d)</td>
<td>16 (100)</td>
</tr>
<tr>
<td>Machine score vs examiner-2 score</td>
<td>0.629</td>
<td>.009(^b)</td>
<td>16 (100)</td>
</tr>
<tr>
<td>Examiner 1 score vs examiner-2 score</td>
<td>0.855</td>
<td>&lt;.001(^e)</td>
<td>16 (100)</td>
</tr>
</tbody>
</table>

\(^a\)CAD: computer-aided design.  
\(^b\)RPD: removable partial denture.
Figure 9. Spearman correlation analysis of the virtual and interprofessional objective structured clinical examination scores. (A) For the fixed prosthodontics topic, the positive correlations between the scores of the tooth preparation station and the CAD wax pattern station were significant. (B) For the clinical diagnostics topic, the positive correlations between the virtual standardized patient machine score and the examiners’ scores were significant. CAD: computer-aided design.

To explore the relationship between the objective and subjective evaluations, correlation analysis was conducted between the viOSCE scores and the RIPLS scores as well as between the viOSCE scores and the mutual evaluation scale scores. Insignificant negative correlations were noted between the subjective evaluation scores presented by RIPLS and viOSCE. Similarly, the correlation of the mutual evaluation scale score with the viOSCE scores was not significant. The SD of the scores on the mutual evaluation scale showed a decreasing trend among students with higher viOSCE scores and those with lower scores, but an increasing trend was observed among those with median scores (Table 8 and Figure 10).
Table 8. Spearman correlation analysis between the subjective and objective evaluations presented by the Readiness for Interprofessional Learning Scale (RIPLS) and mutual evaluation scale.

<table>
<thead>
<tr>
<th>Correlation coefficient</th>
<th>P value (2-tailed)</th>
<th>Participants (n=16), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The examiner panel scores of viOSCE(^a) vs the intragroup mean score of RIPLS (before the test)</td>
<td>−0.272</td>
<td>.15</td>
</tr>
<tr>
<td>The examiner panel scores of viOSCE vs the intragroup mean score of RIPLS (after the test)</td>
<td>−0.302</td>
<td>.13</td>
</tr>
<tr>
<td>The examiner panel scores of viOSCE vs the intragroup mean score of the mutual evaluation scale</td>
<td>−0.038</td>
<td>.44</td>
</tr>
</tbody>
</table>

\(^a\)viOSCE: virtual and interprofessional objective structured clinical examination.

Figure 10. Correlation analysis of the viOSCE scores and the mutual evaluation scale. (A) Correlation analysis of the viOSCE scores and the SD of the scores for each item on the mutual evaluation scale. (B) Correlation analysis of the viOSCE scores and the SD of the mean scores on the mutual evaluation scale. viOSCE: virtual and interprofessional objective structured clinical examination.

In the one-on-one interviews, 29 (91%) of the 32 students approved of the effectiveness of viOSCE and wanted to use it to assess their IPCP ability in the graduation examination. At the fixed prosthodontics station, 56% (9/16) of the dental technology students complained about the lack of lingual space prepared by their partners at the tooth preparation station, which made it difficult to design crown wax patterns, and the corresponding dental students reported not being aware of the condition before viOSCE. At the removable prosthodontics station, almost all the dental students (15/16, 94%) reported that the advice given by the dental technology students was effective in helping them complete the RPD design and considered their
design practice to be insufficient. In contrast, the dental technology students reported that helping the dental students complete the RPD design made them feel satisfied with their professional competence and felt that they were truly part of the team during the collaboration. At the clinical diagnosis station, the dental students felt that their clinical practice experience was not sufficient, especially when the dental technology students could provide a diagnostic plan faster than themselves.

In terms of positive feedback, the students believed that viOSCE promoted the friendship between themselves and their partners, helped them realize the continuity and relevance between their own work and the work of their partners, and enabled them to acquire a deep understanding of IPCP. The negative feedback mainly focused on their lack of clinical knowledge, inadequate preparation, and long waiting time at some stations.

**Discussion**

**Principal Findings**

The IPCP results of dentists and dental technicians reflect the quality of their IPE, skill training, and clinical experience. The results contribute to the much-needed IPE assessment literature and suggest that teamwork skills can be improved by IPCP and effectively assessed using this new evaluation scale. We used a modified Delphi process in this study. This is in accordance with Simmons et al [27], who found that the modified Delphi process is an effective tool to obtain consensus among professionals for the foundational work required. In addition, our study demonstrated the effectiveness of iOSCE in asynchronous and synchronous collaboration scenarios, while providing a methodological reference for developing a new iOSCE for dental health care professionals. As the collaboration scenario between the dentist and the dental technician may be both asynchronously applied through prescriptions and synchronously conducted in chairside discussions [61-63], it was deemed appropriate for the viOSCE framework to consider both synchronous and asynchronous scenarios.

The viOSCE scores in this study also reflect the effectiveness of the framework design. From the viOSCE examiner scores in the fixed prosthodontics section, a significant positive correlation between the scores of the tooth preparation station and the CAD wax pattern station was evident. This finding is consistent with the actual clinical asynchronous delivery scenario, where the dentist’s preparation largely determines the quality of the dental technician’s crown wax pattern. Qualitative evaluations extracted from the one-on-one interviews also supported this result. For the removable prosthodontics section, the negative correlation between the scores of the RPD design station and the CAD wax pattern station was not statistically significant, which might be owing to the fact that more than half of the groups (15/16, 94%) worked collaboratively to modify the RPD design to possibly compensate for the lack of training, which is consistent with the findings about dentists’ inadequate competence in RPD design reported in other studies [64,65].

As OSCE is essentially a simulated scenario-based examination, the use of virtual technology to build simulated scenarios has become an important direction for OSCE-related studies, especially in the field of dental education [43]. The COVID-19 pandemic has further contributed to dental educators’ interest in this area, as dental clinical practice typically occurs in a virus-laden aerosolized environment [66]. Therefore, providing a safe and robust learning environment in the simulation clinic is also critical to help students compensate for lost educational time. The virtual technologies used to construct the simulated clinical environment in this study include VSP and CAD. Previously, Janda et al [67] developed a virtual patient as a supplement to standard instruction in the diagnosis and treatment planning of periodontal disease. However, it could not fully understand complex or ambiguous questions, and the students felt frustrated during the practice [67].

Tanzawa et al [68] developed a robot patient that could reproduce an authentic clinical situation and introduced it into OSCE. However, the dialogue recognition of the robot patients was prespecified; the robot was unable to identify subjective patient descriptions or the dentist’s interrogation intention and could not support intraoral or extraoral examinations to obtain diagnostic evidence [68]. To fill these gaps, our study used VSP with intention recognition and haptic feedback to construct virtual dental clinical practice and diagnosis scenarios more realistically. As the diagnostic evidence collected by students through interrogation, inspection, and palpation was automatically summarized for the final differential diagnosis, and omissions in the examination process eventually led to a misdiagnosis, the system simulated a high-fidelity clinical environment. In addition, the results showed that 1 (6%) of the 16 student groups misdiagnosed their VSP because of incomplete interrogation and palpation. The correlations between the scores of the 2 examiners and the machine scores were statistically significant, thus confirming the robustness of the high-fidelity simulation scenarios constructed by the VSP and the machine scores. On the basis of these results, the use of VSP should be expanded and integrated into daily teaching to give students more opportunities for clinical practice training.

Consistent with the results of previous OMEDT studies [56], the use of CAD technology in viOSCE significantly reduced the time spent at each station for the dental technology students. Some dental technology students complained about the slowness of the CAD program. Upon further investigation, it was found that they imported both impressions at the same time. In dental laboratory practice, dental technicians usually import the impressions separately to prevent computational issues. This finding exposes the lack of virtual dental laboratory practice skills in teaching, which needs to be addressed.

The results showed that the teamwork and collaboration subscale scores were significantly increased at the end of the study (F=0.004), suggesting that viOSCE can improve students’ teamwork skills. The increase in the other 3 subscale scores, although not statistically significant, can be explained by the choice of timing of viOSCE. The optimal time to expose medical students to IPE is still subject to debate [18]. viOSCE, as a clinical IPCP intervention introduced during the clinical year, had no significant effect on the promotion of negative or positive
identity or roles and responsibilities. This finding may be due to the fact that the students’ professional cognition had been stereotyped at this time, making it difficult to effect significant changes through IPE or IPCP intervention. This conclusion is supported by a previous study [69].

The results of the mutual evaluation scale showed statistically significant difference in participant motivation between the 2 professional groups, which could be explained by the results of the roles and responsibilities subscale. Of the 16 dental technology students, 4 (25%) expressed that they would not practice as dental technicians in the future because they wanted to choose other careers. The differences in the scores of the other items were not statistically significant, thus showing the effectiveness of viOSCE in the development of teamwork spirit. This result confirms that the OSCE design is well suited as a final evaluation of IPE and IPCP. In addition, the average score of each item of viOSCE was >60, indicating that the students were satisfied with the design and operation of viOSCE. The differences in scores between the 2 types of professionals were not statistically significant, except at the removable prosthodontics station, which was probably caused by the dental technology students’ unfamiliarity with the CAD program.

Overall, the internal consistency of all subjective evaluations was acceptable, and the results met expectations. Interesting observations were also made regarding the correlation between the subjective and objective evaluations. The SD of the scores on the mutual evaluation scale showed a decreasing trend among the dental and dental technology students with higher viOSCE scores and those with lower scores, but an increasing trend in the median score was observed. Although this trend was not statistically significant due to sample size limitations, this early finding provides data support for a summary of clinical experience published previously by Preston [70], who reported the intensity of the relationship between dentists and dental technicians is determined by the difference in their professional skills. If the professional skills of both parties are high, there will be few problems in their cooperative relationship. The more discriminating and demanding the technician or dentist becomes, the more the relationship is strained when either fails to perform up to the other’s standards. This result suggests that in the study of IPE and IPCP for dentists and dental technicians, it is not sufficient to explore the improvement of the traditional assessment dimensions such as team collaboration skills and identity. The final quality of the output must be included in the assessment dimension. This also reaffirms the effectiveness of viOSCE as an objective, quantitative evaluation tool for IPE and IPCP.

Limitations and Future Studies
The main limitation of our study is the small convenience sample of participating students, which could have led to self-selection bias. The sample size should be expanded in the future to obtain more data and to further verify the robustness of the viOSCE framework. In addition, whether viOSCE should be made a part of the large and more complete OSCE to test the ability of students to meet undergraduate graduation requirements will also be the focus of our next study. Moreover, the independent application of the novel VSP in the education of dental students is an interesting topic that will be explored in the next step of this study.

Recommendations
On the basis of our results, we provide the following recommendations:

1. All dental health professionals should be educated to deliver patient-centered care as members of an interdisciplinary team [16].
2. IPE intervention–related skills should be introduced as preclinical skills.
3. The cooperation of the dental care team is complex, and the training for improving the cooperation ability of the dental care team should include both subjective and objective assessments.
4. viOSCE and scale assessment should be introduced for the assessment of IPE and IPCP at the clinical stage of training.

Conclusions
In this study, a novel viOSCE framework was developed and piloted. Data based on subjective evaluation scales and objective examiner scores were collected and analyzed, confirming the effectiveness of viOSCE as an objective evaluation tool for IPE and IPCP. The experimental design should be expanded to include more randomly selected students with a scientifically determined sample size to further develop studies focused on IPE and IPCP in dentistry and dental technology, ultimately promoting quality in dental clinical practice.

Acknowledgments
This study was funded by a major project of Chongqing Higher Education and Teaching Reform of the People’s Republic of China (reference number 201019 to JS and reference number 193070 to MP) and the Teaching Reformation Fund of the College of Stomatology, Chongqing Medical University (KQJ202105 to YD and KQJ202003 to MP). All authors thank Beijing Unidraw Virtual Reality Technology Research Institute Co Ltd for assisting in the pilot study and the maintenance of the virtual standardized patient and Objective Manipulative Skill Examination of Dental Technicians systems.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.
Authors' Contributions

MP was the principal investigator who organized the group of experts to develop the framework of the virtual and interprofessional objective structured clinical examination (viOSCE); designed the validity experiment; collected and analyzed the data; and drafted the manuscript. YD organized the viOSCE pilot study and assisted with fundraising and distribution, expert panel recruitment, and data collection. XZ and JW led an engineering team from Beijing Unidraw Virtual Reality Technology Research Institute Co Ltd to complete the development of the virtual standardized patient and Objective Manipulative Skill Examination of Dental Technicians systems and assisted in the maintenance of the virtual standardized patient and Objective Manipulative Skill Examination of Dental Technicians systems during the viOSCE pilot study. Li Jiang assisted in the preparation of clinical cases and equipment related to viOSCE, checked all the details, and was responsible for maintaining order at the viOSCE facility. PJ and JS supervised and directed the project. Lin Jiang supervised the advancement of the project and assisted in recruiting the participating students. All authors approved the manuscript.

Conflicts of Interest

None declared.

References


61. Mykhaylyuk N, Mykhaylyuk B, Dias NS, Blatz MB. Interdisciplinary esthetic restorative dentistry: the digital way. Compendium 2021 Nov;42(10) [FREE Full text]


Abbreviations

CAD: computer-aided design
IPCP: interprofessional collaborative practice
IPE: interprofessional education
OMEDT: Objective Manipulative Skill Examination of Dental Technicians
OSCE: objective structured clinical examination
PI: principal investigator
RIPLS: Readiness for Interprofessional Learning Scale
RPD: removable partial denture
vOSCE: virtual and interprofessional objective structured clinical examination
vOSCE: virtual objective structured clinical examination
VSP: virtual standardized patient

©MengWei Pang, YanLing Dong, XiaoHan Zhao, JiaWu Wan, Li Jiang, JinLin Song, Ping Ji, Lin Jiang. Originally published in JMIR Formative Research (https://formative.jmir.org), 17.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

https://formative.jmir.org/2024/1/e44653

JMIR Form Res 2024 | vol. 8 | e44653 | p.504https://formative.jmir.org/2024/1/e44653

Edited by A Mavragani; submitted 28.11.22; peer-reviewed by C Fonseca, Y Zhou, N Ab Malik; comments to author 16.03.23; revised version received 26.04.23; accepted 13.12.23; published 17.01.24.

Please cite as:

Pang M, Dong Y, Zhao X, Wan J, Jiang L, Song J, Ji P, Jiang L. Virtual and Interprofessional Objective Structured Clinical Examination in Dentistry and Dental Technology: Development and User Evaluations
JMIR Form Res 2024;8:e44653
URL: https://formative.jmir.org/2024/1/e44653
doi:10.2196/44653
PMID:38231556
Use of Machine Learning Tools in Evidence Synthesis of Tobacco Use Among Sexual and Gender Diverse Populations: Algorithm Development and Validation

Shaoying Ma\textsuperscript{1*}, PhD; Shuning Jiang\textsuperscript{2}, BS; Olivia Yang\textsuperscript{2}; Xuanzhi Zhang\textsuperscript{2}, BS; Yu Fu\textsuperscript{2}, BS; Yusen Zhang\textsuperscript{2}, BS; Aadeeba Kaareen\textsuperscript{1}, BSoCSci; Meng Ling\textsuperscript{2}, PhD; Jian Chen\textsuperscript{2}, PhD; Ce Shang\textsuperscript{1}, PhD

\textsuperscript{1}Center for Tobacco Research, The Ohio State University Comprehensive Cancer Center, Columbus, OH, United States
\textsuperscript{2}Department of Computer Science and Engineering, The Ohio State University, Columbus, OH, United States
\textsuperscript{*}these authors contributed equally

Corresponding Author:
Shaoying Ma, PhD
Center for Tobacco Research
The Ohio State University Comprehensive Cancer Center
3650 Olentangy River Road
1st Floor, Suite 110
Columbus, OH, 43214
United States
Phone: 1 6148976063
Email: shaoying.ma@osumc.edu

Abstract

\textbf{Background:} From 2016 to 2021, the volume of peer-reviewed publications related to tobacco has experienced a significant increase. This presents a considerable challenge in efficiently summarizing, synthesizing, and disseminating research findings, especially when it comes to addressing specific target populations, such as the LGBTQ+ (lesbian, gay, bisexual, transgender, queer, intersex, asexual, Two Spirit, and other persons who identify as part of this community) populations.

\textbf{Objective:} In order to expedite evidence synthesis and research gap discoveries, this pilot study has the following three aims: (1) to compile a specialized semantic database for tobacco policy research to extract information from journal article abstracts, (2) to develop natural language processing (NLP) algorithms that comprehend the literature on nicotine and tobacco product use among sexual and gender diverse populations, and (3) to compare the discoveries of the NLP algorithms with an ongoing systematic review of tobacco policy research among LGBTQ+ populations.

\textbf{Methods:} We built a tobacco research domain–specific semantic database using data from 2993 paper abstracts from 4 leading tobacco-specific journals, with enrichment from other publicly available sources. We then trained an NLP model to extract named entities after learning patterns and relationships between words and their context in text, which further enriched the semantic database. Using this iterative process, we extracted and assessed studies relevant to LGBTQ+ tobacco control issues, further comparing our findings with an ongoing systematic review that also focuses on evidence synthesis for this demographic group.

\textbf{Results:} In total, 33 studies were identified as relevant to sexual and gender diverse individuals’ nicotine and tobacco product use. Consistent with the ongoing systematic review, the NLP results showed that there is a scarcity of studies assessing policy impact on this demographic using causal inference methods. In addition, the literature is dominated by US data. We found that the product drawing the most attention in the body of existing research is cigarettes or cigarette smoking and that the number of studies of various age groups is almost evenly distributed between youth or young adults and adults, consistent with the research needs identified by the US health agencies.

\textbf{Conclusions:} Our pilot study serves as a compelling demonstration of the capabilities of NLP tools in expediting the processes of evidence synthesis and the identification of research gaps. While future research is needed to statistically test the NLP tool’s performance, there is potential for NLP tools to fundamentally transform the approach to evidence synthesis.

(JMIR Form Res 2024;8:e49031) doi:10.2196/49031
KEYWORDS
machine learning; natural language processing; tobacco control; sexual and gender diverse populations; lesbian; gay; bisexual; transgender; queer; LGBTQ+; evidence synthesis

Introduction
The use of nicotine or tobacco products is a leading preventable cause of cancer, heart diseases, and lung diseases in the United States [1], with cigarette smoking alone responsible for the death of half a million Americans each year [2]. Notably, sexual and gender diverse individuals, often referred to as the LGBTQ+ (lesbian, gay, bisexual, transgender, queer, intersex, asexual, Two Spirit, and other persons who identify as part of this community) populations, are particularly vulnerable to nicotine and tobacco product use [3]. Both the National Cancer Institute and the Centers for Disease Control and Prevention have recognized the LGBTQ+ populations as a critical target in their efforts to combat tobacco use disparities [4-10].

In response to the pressing need for tobacco control and the rapidly evolving landscape of the tobacco market, the National Institutes of Health (NIH) and other health foundations, including the American Cancer Society, have made substantial investments in tobacco control research and tobacco regulatory science [11,12]. According to our calculations using data from the NIH era reporter, funding for tobacco research has shown a remarkable increase, growing from US $7.7 billion in 2016 to US $11.2 billion in 2021 (Multimedia Appendix 1 [13]). Consequently, the volume of peer-reviewed publications related to tobacco has experienced a significant increase. This presents a considerable challenge in efficiently summarizing, synthesizing, and disseminating research findings, especially when it comes to addressing specific target populations, such as the LGBTQ+ populations.

One promising pathway to rapidly assessing the expanding body of literature is the use of natural language processing (NLP) models. NLP is dedicated to deciphering and comprehending how computers interpret human language, equipping them to analyze extensive data sets of natural language [14-16]. While NLP tools have garnered considerable recognition in biomedical research [4-10], aiding in tasks such as disease surveillance (eg, COVID-19) and diagnosing using medical records [17-23], their potential to expedite near real-time synthesis of evidence in tobacco control research remains untapped [24].

Another gap in existing NLP tools is the lack of applications in synthesizing social science research and modeling. A noteworthy example in the domain of tobacco research is the evaluation of the effectiveness of tobacco control policies, which are often assessed using complex statistical modelling and large-scale survey data. These methods demand a specialized semantic database for labelling studies and interpreting results. However, to the best of knowledge, such a semantic database has not been developed yet. Considering that policy interventions at federal, state, and local levels are designed to reach a large number of populations, the lack of a database to facilitate NLP applications may significantly undermine evidence synthesis and thereby the timely adoption of effective policies [25].

Furthermore, in light of the calls from entities such as the NIH and other health agencies to address tobacco use disparities within priority populations, including LGBTQ+ populations, the development NLP tools to aid in the discovery of effective policies tailored to these special populations remains uncharted territory [26-31]. There is an urgent demand for the development of NLP tools (eg, semantic database, NLP algorithms) in tobacco research that have the abilities to synthesize evidence in social science and assist in research gap discovery for priority populations.

In this pilot study, we aimed to achieve the following goals to address the identified research and application gaps: (1) compile a specialized semantic database for tobacco policy research to extract information from journal article abstracts, (2) develop NLP algorithms that comprehend the literature on nicotine and tobacco product use among sexual and gender diverse populations, and (3) compare the discoveries of the NLP algorithms with an ongoing systematic review of tobacco policy research among LGBTQ+ populations [32]. While this pilot study does not fully address the gaps by developing a comprehensive evidence synthesis or discovery tool for tobacco research, the outcomes may pave the road for future tools that can achieve this goal. Our vision is that NLP tools may be able to assist academic scholars and policy makers in prescribing public health policies, such as tobacco control policies, and addressing public health needs, such as reducing health disparities.

Methods
Development of a Tobacco Research Domain–Specific Semantic Database
Overview
To generate a tobacco research domain-specific semantic database, we used an iterative process that combines expert opinions and the reading of tobacco research papers in 4 leading tobacco journals (Tobacco Control, Nicotine and Tobacco Research, Tobacco Induced Diseases, and Tobacco Prevention and Cessation). The main categories of keywords were the follows: (1) tobacco use behaviors, prevalence, and outcomes; (2) population characteristics; (3) geographic locations; (4) method and inference; (5) policy; (6) tobacco products; (7) relation statement; and (8) tobacco characteristics. Under each main category, there were one or more subcategories, and each subcategory contained a list of named entities. Table 1 presents the categories of named entities in a domain-specific semantic database that were used for training and improving a language model for tobacco research on sexual and gender diverse populations. These categories are based on journal articles’ keywords, further guided by existing literature on how to use NLP methods to synthesize public health evidence [25,33]. These categories are important components of a study, encompassing measures, methods, results, conclusions, and hypothesis testing.
Table 1. Main categories and subcategories of named entities.

<table>
<thead>
<tr>
<th>Main categories</th>
<th>Subcategories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobacco use behavioral outcomes</td>
<td>• Tobacco cessation</td>
</tr>
<tr>
<td></td>
<td>• Exposure to tobacco-related or antitobacco content, or exposure to secondhand or thirdhand smoking</td>
</tr>
<tr>
<td></td>
<td>• Health and disease</td>
</tr>
<tr>
<td></td>
<td>• Perception and belief</td>
</tr>
<tr>
<td></td>
<td>• Tobacco use prevalence</td>
</tr>
<tr>
<td></td>
<td>• Time period</td>
</tr>
<tr>
<td>Population characteristics</td>
<td>• Age groups</td>
</tr>
<tr>
<td></td>
<td>• Sex</td>
</tr>
<tr>
<td></td>
<td>• Sexual and gender diverse populations</td>
</tr>
<tr>
<td></td>
<td>• Racial and ethnic minoritized groups</td>
</tr>
<tr>
<td></td>
<td>• Socioeconomic status</td>
</tr>
<tr>
<td>Geographic locations</td>
<td>• Countries, states, provinces, or cities</td>
</tr>
<tr>
<td>Method and inference</td>
<td>• Data</td>
</tr>
<tr>
<td></td>
<td>• Methodology</td>
</tr>
<tr>
<td></td>
<td>• Statistics</td>
</tr>
<tr>
<td>Policy</td>
<td>• Marketing</td>
</tr>
<tr>
<td></td>
<td>• Law, policy, and regulation</td>
</tr>
<tr>
<td></td>
<td>• Regulation body</td>
</tr>
<tr>
<td></td>
<td>• Treatment</td>
</tr>
<tr>
<td>Tobacco products</td>
<td>• Combustible tobacco products</td>
</tr>
<tr>
<td></td>
<td>• Noncombustible tobacco products</td>
</tr>
<tr>
<td>Relation statement</td>
<td>• Relation terms</td>
</tr>
<tr>
<td>Tobacco characteristics</td>
<td>• Chemical</td>
</tr>
<tr>
<td></td>
<td>• Flavor</td>
</tr>
</tbody>
</table>

**Journal Selection**

We chose 4 peer-reviewed tobacco-specific multidisciplinary journals, namely, *Tobacco Control*, *Nicotine and Tobacco Research*, *Tobacco Induced Diseases*, and *Tobacco Prevention and Cessation*, to extract articles and compile keywords at the initial stage. The first 2 are among the journals that have the highest impact factors in addiction research; in 2022, *Tobacco Control* had an impact factor of 5.2 and a 5-year impact factor of 5.7 [34], and *Nicotine and Tobacco Research* had an impact factor of 4.7 and a 5-year impact factor of 4.2 [35]. *Tobacco Induced Diseases* [36] and *Tobacco Prevention and Cessation* [37] are 2 other peer-reviewed journals that specifically publish research on nicotine and tobacco products but are not as highly ranked as the other 2 journals. The textual data from the 4 peer-reviewed journal articles contained a total of 2993 abstracts from published papers from 2015 to early 2021.

While the 2993 articles extracted from these journals do not represent the full body of tobacco research, they cover a significant share of tobacco studies and integrate evidence across the 5 translational research stages: basic research, preclinical research, clinical research, clinical implementation, and public health. These journals also ask authors to specify how the research reported contributes to tobacco control objectives, which have policy implications. Alternatively, a random sampling from PubMed searches using tobacco related terms may not yield studies that are necessarily translational in nature. Therefore, we focused on the articles published in the 4 journals in our study.

**Iterative Process to Expand Terms (Named Entities) in the Database**

The general process included the following iterative steps: (1) to generate initial annotation data, we first compiled key terms from extracted articles and allocated key terms to categories using group discussions; (2) we enriched the database using various sources and group discussions (more specific descriptions below); (3) we fine-tuned the spaCy en_core_web_lg model with the initial annotation and following iterative versions of data (the en_core_web_lg model is a pretrained large language model that can extract multiple general named entities); (4) we expanded the list of named entities to include more keywords of similar meanings using SeedNER [38,39], that is, a small set of initial labeled examples or patterns that was used as a starting point for training a model; (5) we searched the occurrence of each keyword in the 2993 paper abstracts and kept those with high frequency; (6) during this process, named entities that were too generic to yield meaningful relations were removed from the database; and (7) we repeated steps 3 to 6 until the set of entities reached our satisfaction during group discussions.
Specific approaches were used for conducting step 2. For categories including “tobacco use behavioral outcomes,” “tobacco products,” and “tobacco characteristics,” the iterative process involved four steps: (1) discussions to determine whether to include newly identified key terms and how to allocate them into additional subcategories (Table 1); (2) using a named entity recognition (NER) model to extract named entities from 2993 paper abstracts from the 4 specific journals; (3) randomly sampling and reviewing the output of the NER model, correcting identified errors, and adding missed NERs; and 4) repeating steps 1 to 3 until we were satisfied with the model output.

The categories “population characteristics,” “geographic locations,” and “relation terms” are commonly used concepts in real life and not specific to tobacco control. We used Google searches, Wikipedia, and WordNet to enrich the key terms. In addition, for the “method and inference” category, we used the glossary of an econometrics methodology textbook by Cameron and Trividi to enrich the terms [40]. This textbook is widely used in economics and social science and its glossary should provide sufficient terms for this category.

For the “policy” category, we drew named entities from 2 sources that comprehensively summarize available tobacco control policies in the regulatory space. The first source was a peer-reviewed journal article by McDaniel et al [41] that conducted an intensive policy scan of all possible regulations that can contribute to tobacco endgame. The second source was the World Health Organization’s report on the global progress in implementing tobacco control policies, as recommended by the World Health Organization’s Framework Convention on Tobacco Control [42], which is the largest public health treaty signed by 182 countries and prescribes a comprehensive set of tobacco control policies. These policies are classified into 5 groups: M (monitor tobacco use and prevention policies), P (protect people from tobacco smoke), O (offer help to quit tobacco use), W (warn about the dangers of tobacco), E (enforce bans on tobacco advertising, promotion, and sponsorship), and R (raise taxes on tobacco) [42]. These sources cover policy key terms related to both national and international contexts and together create the most comprehensive policy terms to our knowledge.

Development of NLP Algorithms That Comprehend the Literature on Nicotine and Tobacco Product Use Among Sexual and Gender Diverse Populations

We used RoBERTa, an optimized BERT (bidirectional encoder representations from transformers-based) language model [43], to perform NER tasks. BERT is a state-of-the-art language model that excels at tasks such as sentiment analysis and text summarization. By learning patterns and relationships between words and their context in text, BERT can extract named entities that it has learned during training and potentially discover new ones.

We developed an NER model based on RoBERTa using the Python (Python Software Foundation) programming language and the spaCy library [44]. We began by defining 36 labels of categories (main and subcategories; Table 1) and extracting 1582 named entities using the existing NER model RoBERTa. Next, those named entities were used to tag abstracts and create a training set, using the annotation tool Prodigy [45]. A subset of the abstracts with labeled named entities was reviewed by 2 domain experts to identify key terms that were missing in our semantic database, which were added to the lists of named entities.

The RoBERTa model was then updated based on the richer database and further trained for a maximum of 20,000 steps, with early stopping implemented if no improvement was observed for 1600 consecutive steps. With a series of iterations, we used the updated RoBERTa model to assess the 2993 abstracts and labeled them with the categories.

When identifying studies related to LGBTQ+ populations, it is important to understand that this community is heterogeneous [46,47]. Given that LGBTQ+ key terms are included in the “population characteristics” categories, we were able to identify LGBTQ+ populations based on categorization. There were 111 LGBTQ+-related named entities in our database.

Comparison of the Discoveries of the NLP Algorithms

With an Ongoing Systematic Review of Tobacco Policy Research Among LGBTQ+ Populations

Ideally, we would like to compare the results from our tools with those from systematic reviews and meta-analyses of studies related to tobacco control issues among LGBTQ+ populations. Systematic reviews and meta-analyses are state-of-the-art evidence synthesis methods that can provide the ground truth [48-50]. While we are currently conducting a separate systematic review of the effectiveness of tobacco control policies among LGBTQ+ populations, this review has not been finalized yet [32]. Nonetheless, the ongoing systematic review does provide some data points for comparisons, including the number of studies extracted from the 4 journals and presence of policy assessment. Therefore, we conducted comparisons of these 2 domains.

Ethical Considerations

This study does not involve human subjects, as it synthesizes data from research articles published at peer-reviewed journals. The Ohio State University Institutional Review Board has determined that it contains no human subjects and thus no further review is needed (study number: 2021E0776).

Results

In total, we identified 33 articles relevant to sexual and gender diverse populations from the 2993 abstracts. Our trained model successfully extracted 773 named entities (181 unique named entities) from the 33 paper abstracts to describe the themes of these articles. Among the 773 extracted named entities, 688 were already learned by the model during training, while 70 were new time- or age-related words (eg, 18 years, 2013), 9 were new statistical terms (eg, N=20), and 6 were newly discovered and labeled within other categories. We did not observe any newly discovered policy-related terms.

In Figures 1-3, we present the hierarchy of named entities extracted from abstracts in published papers that studied nicotine...
or tobacco product use among sexual and gender diverse individuals. Each number on the right is the frequency of the corresponding named entity by paper abstract. Named entities with the same color belong to the same main category.

**Figure 1.** Hierarchy and frequency counts of named entities extracted from published research in tobacco-specific journals from 2015 to early 2021 in 4 main categories: tobacco use, products, characteristics, and relation statement. Numbers represent the frequency of the corresponding named entity by paper abstract.
Figure 2. Hierarchy and frequency counts of named entities extracted from published research in tobacco-specific journals from 2015 to early 2021 in the main category of population characteristics. Numbers represent the frequency of the corresponding named entity by paper abstract.
According to our tool, among the 33 tobacco studies related to LGBTQ+ populations, the most frequent use outcomes were “cigarette smoking” (n=17), “substance use” (n=16), “prevalence” (n=16), and “risk” perception (n=14). Also, for these populations, “cigarettes” (n=15) were the most frequently mentioned combustible tobacco product and “e-cigarettes are” (n=8) was the most frequently mentioned noncombustible tobacco product. In addition, for tobacco characteristics, “alcohol” (n=7) and “nicotine” (n=5) were the most mentioned attributes among LGBTQ+ tobacco research papers.

The relation statement findings suggest that a majority of the studies examined “comparison” (n=26), “association” (n=23), and “correlation” (n=6). We found no studies that explicitly used the term “causal” or “causality” in the studies.
The population characteristics mentioned in the studies illustrated that among socioeconomic status terms, the most frequently included were “demographics” (n=8) and “SES factors” (n=8). Among sex and sexual and gender minority terms, the most frequent ones were “bisexual” (n=21), “lesbian” (n=19), and “gay” (n=19). Among racial and ethnic minority group terms, the most frequent ones were “minority groups” (n=8) and “Race/ethnicity” (n=6). For age group terms, the terms included “adult” (n=14), “young adult” (n=11), “adolescent” (n=5), “students” (n=3), and “adolescents and young adults” (n=2).

The policy category showed that in these studies, the most mentioned term was “intervention” (n=16). In addition, while the general term “tobacco control” was mentioned in 6 studies, only 1 study contained any specific policy term (“smoke free air law”). As such, there was a significant gap in policy research among the published articles in the 4 leading tobacco journals between 2015 and early 2021, since only 1 study mentions specific policies when it comes to tobacco research among the LGBTQ+ populations. The statistics and methodology terms further indicated that the most used terms included “survey” (n=18) and “logistic regression” (n=10), and relatively fewer studies mentioned terms related to causal inferences, such as “experimental research” (n=4), “randomization” (n=3), and “clinical trial” (n=1). The studies mentioning “US” also dominated in the numbers, with 12 studies in total. Several studies that assessed countries with multilevel governing levels, such as Canada and the United States, also appeared to have mentioned “state,” “city,” and “province,” suggesting that attention was paid to these defined areas.

We next compared our results using the NLP tools with our ongoing systematic review. Similar to the conclusions of the ongoing systematic review, we found very few studies that yielded specific policy recommendations. This finding was further corroborated by the lack of causal inference methods labeled by the NLP tool. While our NLP tool cannot replace systematic reviews just yet, it does show potential to complement the existing methods and requires less human supervision (systematic reviews usually require at least 2 human coders).

**Discussion**

This pilot study builds a semantic database dedicated to tobacco research and developed NLP algorithms to automatically identify, extract, and summarize textual data from published tobacco studies. We further demonstrated a user case wherein we assessed LGBTQ+ tobacco research by labeling key components of a tobacco study: tobacco use outcomes, tobacco characteristics, population characteristics, geographic locations, method and inference, and policy relevance.

It is worth noting that the components we categorized, such as “method and inference,” align with the typical sections found in scientific articles in social science, including measures, methods, results, conclusions, and hypothesis testing. As a result, our tool extracts text segments that are frequently assessed in evidence synthesis, thereby showing the potential of using NLP tools to enhance systematic reviews and facilitate meta-analyses.

Additionally, we leveraged the NLP algorithms we created to identify gaps in tobacco research concerning the LGBTQ+ populations and concluded that there is a scarcity of studies assessing policy impacts on this demographic using causal inference methods. This finding is consistent with our ongoing systematic review [32], highlighting how NLPs have the capacity to aid in both evidence synthesis and research gap discoveries. This, in turn, has the potential to streamline research efforts, reduce labor costs, and influence the trajectories of future research directions [51,52].

Using the NLP tool, we further found some interesting patterns in tobacco research involving LGBTQ+ populations. It appears that the product drawing the most attention in the field is cigarettes or cigarette smoking and that the number of studies of various age groups is almost evenly distributed between youth or young adults and adults. Moreover, the existing evidence body is dominated by studies coming from the United States. These patterns are consistent with the research needs to reduce cigarette smoking among LGBTQ+ populations in the United States, where 16.1% of LGBTQ+ adults and 17.4% of LGBTQ+ high schooler students smoke cigarettes—this is 4% to 6% higher than their heterosexual counterparts [53,54]. Therefore, our findings align with the ongoing research needs and the financial investments made by the US health agencies like the NIH, thereby bolstering the confidence in the NLP tool that we developed.

Finally, while the semantic database and language model in this pilot study are designed to extract and summarize key components of tobacco research, many of the terms and labeling categories are broad and applicable to public health and social science research in general, such as “methods and inference” and “relation terms.” Therefore, our tool has the potential to transform the evidence synthesis paradigm in tobacco control and public health at large by enabling more efficient and effective analyses of large volumes of textual data. Future tool development may extend its reach to other public health domains, fostering the real-time translation of research findings into evidence-based policymaking, thereby contributing significantly to the advancement of public health initiatives.

Our study has several limitations. First, for the development of keywords and the application of the NLP, we focused on 4 peer-reviewed tobacco-specific research journals, which were not representative of the entire tobacco control literature. However, considering the prominence and extensive content covered by these journals, we believe that this selection is unlikely to introduce significant selection bias or result in the omission of crucial keywords. Second, although we used our ongoing systematic review as a benchmark for the qualitative assessment of the results obtained in this pilot study, we did not perform a quantitative comparison of our findings with the ground truth derived from the systematic review. This quantitative evaluation, which might include measures like Cohen kappa, was not conducted because the systematic review has not yet been finalized. Consequently, future research endeavors are required to undertake a thorough quantitative
comparison between the training data and the established ground
truth using statistical testing for a more comprehensive
assessment of the NLP tool’s performance.

Despite the limitations, our pilot study serves as a compelling
demonstration of the capabilities of NLP tools in expediting the
processes of evidence synthesis and the identification of research
gaps. Expanding the scope of this pilot research to encompass
other public health disciplines, extending beyond the realm of
tobacco control, holds the promise of fundamentally
transforming the approach to evidence synthesis. Such expansion
has the potential to play a pivotal role in shaping policy
development across a wide spectrum of public health domains.

Acknowledgments
This study was supported by the President’s Research Excellence (PRE) Accelerator Grant from The Ohio State University
(principal investigator: CS). CS was funded by the National Cancer Institute (R21CA249757). SM was supported by the Pelotonia
Fellowship from The Ohio State University Comprehensive Cancer Center.

During the preparation of this work, the authors used ChatGPT 3.5 in order to check grammar errors and improve language flow.
After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the
publication.

Data Availability
The data sets generated and analyzed during this study are available in the GitHub repository [55].

Authors’ Contributions
CS and SM conceptualized the study. JC, ML, SJ, CS, and SM designed the methodology. SJ and ML were responsible for the
software. CS and JC validated the data. SM and SJ performed the formal analysis. ML, SJ, and SM carried out the investigation.
CS and JC provided resources. ML, OY, XZ, YF, YZ, SJ, and SM performed data curation. SM and SJ wrote the original draft.
SM, SJ, JC, and CS reviewed and edited the manuscript. CS and JC supervised the study. ML and SJ were responsible for project
administration. CS acquired funding. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Tobacco-related funding from the National Institutes of Health (NIH), 2010-2022. Data was obtained from the National Institutes
of Health [13].

References
2. National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. The health
books/NBK179276/ [accessed 2024-01-09]
URL: https://www.cdc.gov/tobacco/health-equity/lgbtq/health-burden.html [accessed 2024-01-09]
events: design of a unified extraction framework based on neural networks. JMIR Med Inform 2022 Jun 07;10(6):e37804
[FREE Full text] [doi: 10.2196/37804] [Medline: 35671070]
5. Elmessiry A, Cooper WO, Catron TF, Karrass J, Zhang Z, Singh MP. Triaging patient complaints: Monte Carlo
cross-validation of six machine learning classifiers. JMIR Med Inform 2017 Jul 31;5(3):e19 [FREE Full text] [doi:
10.2196/medinform.7140] [Medline: 28760726]
semantic textual similarity in the clinical domain: validation study. JMIR Med Inform 2021 Dec 30;9(12):e27386 [FREE
Full text] [doi: 10.2196/27386] [Medline: 34967748]
review. JMIR Ment Health 2020 Apr 22;7(4):e35928 [FREE Full text] [doi: 10.2196/35928] [Medline: 35451984]
sexual and reproductive health in India (SnehAI): instrumental case study. J Med Internet Res 2022 Jan 03;24(1):e29969
[FREE Full text] [doi: 10.2196/29969] [Medline: 34982034]
34. About. Tobacco Control. URL: https://tobaccocontrol.bmj.com/pages/about [accessed 2023-12-01]  
35. About the journal. Nicotine & Tobacco Research. URL: https://academic.oup.com/ntr/pages/About [accessed 2023-12-01]  
36. Aims and scope. Tobacco Induced Diseases. URL: http://www.tobaccoinduceddiseases.org/Aims-and-Scope,386.html [accessed 2023-12-01]  
44. spaCy. URL: https://spacy.io/ [accessed 2024-01-09]  

**Abbreviations**

BERT: bidirectional encoder representations from transformers  
LGBTQ+: lesbian, gay, bisexual, transgender, queer, intersex, asexual, Two Spirit, and other persons who identify as part of this community  
NER: named entity recognition  
NIH: National Institutes of Health  
NLP: natural language processing
Acceptance of Medical Artificial Intelligence in Skin Cancer Screening: Choice-Based Conjoint Survey

Inga Jagemann1, MSc; Ole Wensing1, BSc; Manuel Stegemann1, Prof Dr; Gerrit Hirschfeld1, Prof Dr
School of Business, University of Applied Sciences and Arts Bielefeld, Bielefeld, Germany

Corresponding Author:
Inga Jagemann, MSc
School of Business
University of Applied Sciences and Arts Bielefeld
Interaktion 1
Bielefeld, 33619
Germany
Phone: 49 521106 ext 70508
Email: inga.jagemann@hsbi.de

Abstract

Background: There is great interest in using artificial intelligence (AI) to screen for skin cancer. This is fueled by a rising incidence of skin cancer and an increasing scarcity of trained dermatologists. AI systems capable of identifying melanoma could save lives, enable immediate access to screenings, and reduce unnecessary care and health care costs. While such AI-based systems are useful from a public health perspective, past research has shown that individual patients are very hesitant about being examined by an AI system.

Objective: The aim of this study was two-fold: (1) to determine the relative importance of the provider (in-person physician, physician via teledermatology, AI, personalized AI), costs of screening (free, 10€, 25€, 40€; 1€=US $1.09), and waiting time (immediate, 1 day, 1 week, 4 weeks) as attributes contributing to patients’ choices of a particular mode of skin cancer screening; and (2) to investigate whether sociodemographic characteristics, especially age, were systematically related to participants’ individual choices.

Methods: A choice-based conjoint analysis was used to examine the acceptance of medical AI for a skin cancer screening from the patient’s perspective. Participants responded to 12 choice sets, each containing three screening variants, where each variant was described through the attributes of provider, costs, and waiting time. Furthermore, the impacts of sociodemographic characteristics (age, gender, income, job status, and educational background) on the choices were assessed.

Results: Among the 383 clicks on the survey link, a total of 126 (32.9%) respondents completed the online survey. The conjoint analysis showed that the three attributes had more or less equal importance in contributing to the participants’ choices, with provider being the most important attribute. Inspecting the individual part-worths of conjoint attributes showed that treatment by a physician was the most preferred modality, followed by electronic consultation with a physician and personalized AI; the lowest scores were found for the three AI levels. Concerning the relationship between sociodemographic characteristics and relative importance, only age showed a significant positive association to the importance of the attribute provider (r=0.21, P=.02), in which younger participants put less importance on the provider than older participants. All other correlations were not significant.

Conclusions: This study adds to the growing body of research using choice-based experiments to investigate the acceptance of AI in health contexts. Future studies are needed to explore the reasons why AI is accepted or rejected and whether sociodemographic characteristics are associated with this decision.

(JMIR Form Res 2024;8:e46402) doi:10.2196/46402

KEYWORDS
artificial intelligence; skin cancer screening; choice experiment; melanoma; conjoint analysis, technology acceptance; adoption; technology use; dermatology; skin cancer; oncology; screening; choice based; trust
Introduction

Skin cancers are the most common groups of cancers diagnosed worldwide, with more than 1.5 million new cases estimated in 2020 [1]. Melanoma is the deadliest form of skin cancer. Based on demographic changes, it is estimated that more than 500,000 new cases of melanoma and almost 100,000 deaths from melanoma should be expected worldwide by 2040 [1]. As melanoma case numbers are expected to increase in the future, high-cost treatments will continue to put a strain on the already overburdened health care budgets. To combat the rising mortality rate of melanoma, early detection is critical. Currently, the German national treatment guidelines [2] recommend skin cancer screening as a standardized full-body skin examination performed by dermatologists who have completed specialized training in the early detection of skin cancer. In addition, dermatologists should use dermoscopy to diagnose suspected skin cancer. Given the rising number of cases as well as increasing scarcity of trained dermatologists [3-5], there has been substantial research into the feasibility of artificial intelligence (AI) to augment or replace traditional skin cancer screening regimens [6].

AI describes machines (or computers) that mimic the cognitive functions associated with human thought, such as learning and problem-solving. These systems observe their surroundings and adopt action to reach their targets directly [7]. Further, AI has the ability to learn from images and subsequently provide an image-based diagnosis. Dermatology, as an image-based field of medicine, retains a dominant position in the AI evolution with the ability to classify skin lesions [8].

Research into the technical quality of AI-based skin cancer screening technologies has shown that these systems achieve detection rates that are on par or better than those of highly trained clinicians [9-13]. This highlights the great potential of AI for future skin cancer screening in the general population. As part of apps, AI systems offer immediate access to dermatological screening for all patients with mobile digital devices, enabling health care and treatment to be provided regardless of time and place close to everyday life [6]. Thus, AI systems capable of detecting melanoma and nonmelanoma skin cancer could avoid unnecessary care, reduce health care costs, offer solutions to the increasing scarcity of clinicians, and reduce the waiting times for an appointment and for a diagnosis [3-5]. However, there is a risk that some melanomas will be missed and treatment delayed if the apps incorrectly reassure the user that their lesion is of low risk [14].

Although the technical quality has improved, there is also a growing awareness that patients do not generally accept the use of AI-based systems in health care settings. There is still no consistent definition of technology acceptance in the literature. Terms such as “acceptability,” “acceptance,” and “adoption” are often employed in this context, sometimes interchangeably. Dillon and Morris [15] defined user acceptance “as the demonstrable willingness within a user group to employ IT [information technology] for the tasks it is designed to support.”

Khullar et al [16] conducted an online survey to examine patients’ perspectives about applications of AI in health care, showing that 31% of respondents reported being very uncomfortable and 40.5% were somewhat uncomfortable with receiving a diagnosis from an AI algorithm that was accurate 90% of the time but incapable of explaining its rationale. Longoni et al [17] demonstrated that consumers are very hesitant about being examined by an AI system and consumers’ willingness to pay decreases when an equivalent service is performed by an AI system. Additionally, they concluded that patients’ perceived neglect of uniqueness leads to more resistance to medical AI [17].

Past research has also identified several factors that might impact patients’ preferences to use AI-based health care services. The European Commission [18] interviewed citizens of the 28 member states of the European Union (N=27,900) and concluded that younger participants with a high educational level are more likely to use online health care services. This finding was also replicated in oncology patients, where younger patients indicated higher acceptance of and a greater intention to use digital tools and apps to manage their cancer [19]. The European Commission [18] also found that the opinion on AI strongly depends on exposure to related information and knowledge. This relationship is also supported by a series of experiments showing that resistance to the utilization of medical AI is driven by the subjective difficulty of understanding algorithms [13].

Concerning skin cancer, previous research has shown that patients were generally reluctant to use AI-based systems in the field of dermatology. Snoswell et al [6] examined the consumer preference and willingness to pay for mobile teledermoscopy services in Australia using a discrete-choice experiment (N=199). They found that patients prefer a trained medical professional to be involved in their skin cancer screening and that patients are less willing to pay money for teledermatology [6]. However, Snoswell et al [6] did not take into account sociodemographic factors that may have had an impact on the patients’ decisions. In a multicenter clinical study assessing the performance of automated diagnosis of melanoma with a self-completion questionnaire (N=65), Fink et al [20] found that most patients agreed that computer-assisted diagnoses are trustworthy and may generally improve the diagnostic performance of physicians. However, participants rejected the idea of AI-based systems completely replacing physicians and instead strongly favored hybrid solutions in which diagnoses by a physician are supported by automated systems [20].

To date, only three studies have directly addressed the question of which factors are associated with patients’ preferences regarding AI-based skin cancer screening [21-23]. Ghanie et al [22] studied public interest in teledermatology, which was found to be positively associated with a younger age, higher educational attainment, and higher household income. Chang et al [21] examined sociodemographic differences in teledermatology acceptability with a cross-sectional survey (N=13,996), showing that respondents who were interested in teledermatology were more frequently 18-39 years of age, men, college graduates, and tablet or smartphone users. Similarly, young age, male gender, a previous history of melanoma, and higher educational level were significantly associated with a more positive attitude toward skin cancer-related apps [23]. However, it is unclear whether these results from questionnaires...
can be replicated in choice-based experiments that rely to a lesser degree on introspection and are thus one step closer to actual behavior [24].

As described above, provider and costs for a skin cancer screening have high relevance for the user [6,16,17,20]. For this choice-based conjoint analysis, we further added the attribute waiting time for a diagnosis, because studies have shown a strong negative correlation between patient satisfaction and waiting time [25,26]. AI provides the opportunity to get a skin cancer screening immediately, without any waiting time [3]. Due to the shortage of medical professionals, waiting time for a skin cancer diagnosis is also an important attribute for the user [5,7].

The aim of this study was two-fold based on the following two research questions: (1) How important are the attributes provider, costs for screening, and waiting time for diagnosis for participants’ preference for skin cancer screening? (2) Are sociodemographic characteristics, especially age, systematically related to the relative importance scores of participants to the various attributes?

Figure 1. Example choice set (1€= US $1.09).

Methods

Study Design

This cross-sectional study used a choice-based conjoint analysis to examine the acceptance of medical AI for a skin cancer screening from the user perspective. Conjoint analysis is a quantitative marketing research method that quantifies the value consumers place on the attributes of a product [27]. Respondents are asked to make a choice between 2 or more different choice sets, where each set is described in terms of several predefined attributes, each with different levels. Given a sufficient number of choices per respondent, it is then possible to statistically estimate the importance of each attribute and level for the choice in terms of part-worth utilities. This method offers a behavioral approach and is less susceptible to social desirability and other biases [28].

This study systematically manipulated three attributes (provider, cost, and waiting time) for a hypothetical skin cancer screening. Participants were presented with 12 different choice sets one after another, each consisting of three different modes of skin cancer screenings that were generated by combining different levels of the three attributes (see Figure 1 for an example). The choice sets were generated by the conjointly algorithm using default settings [29].

Survey

Before the survey was conducted, it was tested with the “think-aloud” method by three volunteers to find out if there were any comprehensibility problems. For this purpose, the pretest participants had to speak their thoughts aloud while completing the survey [30].

The questionnaire started with informed consent, where participants were informed about the nature and scope of the survey and about the protection of their data. Before starting the questionnaire, participants completed the consent form and agreed to participate in the anonymous study. The participants then moved on to the choice-based conjoint task, which consisted of 12 different choice sets. The participant’s task for each choice set was to indicate the skin cancer screening that they most prefer (ie, they selected one of the three options as their preferred choice). After responding to the choice sets, participants were asked whether they had undergone a skin cancer screening in the last year and at which type of provider.

Finally, the sociodemographic characteristics (age, gender, education, status, income) were assessed. Finally, the survey asked again whether the data could be used for analysis in anonymized form in case respondents changed their minds during the course of the survey and to filter out people who just wanted to “click through” without seriously answering the questions.

Participants

Recruitment was based on a convenience sample through the social environment; individuals were asked to participate in the open voluntary survey shared with contacts via WhatsApp and Instagram. Standard procedures for conducting and reporting online surveys [31] were followed. Furthermore, conjointly’s default methods were used to identify and bar potential duplicate entries from the same user. Data were collected during the time period of September 29, 2022, through October 20, 2022. The link to the survey was clicked 383 times by unique site visitors. Of these potential respondents, 126 (32.9%) people
filled out the conjoint survey completely and gave their agreement for processing their data. In total, 220 (57.4%) respondents opened the link but did not complete the survey and another 33 (8.6%) people were disqualified from the study because they answered the survey several times. Three people (0.8%) did not give their agreement to process their data and a single respondent (0.3%) was excluded because the survey was answered too quickly. Respondents took an average of 4.7 minutes to complete the survey. Table 1 provides an overview of the respondents’ sociodemographic characteristics. There was a relatively equal proportion of participants identifying as male and female. The average age of the participants was 37.6 years and the median age was 29 years.

Table 1. Sociodemographic characteristics of participants sampled from September 29, 2022, to October 20, 2022 in Germany (N=126).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>58 (46.0)</td>
</tr>
<tr>
<td>Female</td>
<td>67 (53.2)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>Still a student</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>School-leaving qualification</td>
<td>25 (19.8)</td>
</tr>
<tr>
<td>Vocational qualification</td>
<td>34 (27)</td>
</tr>
<tr>
<td>University degree</td>
<td>58 (46)</td>
</tr>
<tr>
<td>Doctorate</td>
<td>2 (1.6)</td>
</tr>
<tr>
<td>Other degree</td>
<td>3 (2.4)</td>
</tr>
<tr>
<td>Not specified</td>
<td>3 (2.4)</td>
</tr>
<tr>
<td><strong>Employment status</strong></td>
<td></td>
</tr>
<tr>
<td>Elementary/high school student</td>
<td>2 (1.6)</td>
</tr>
<tr>
<td>University student</td>
<td>29 (23)</td>
</tr>
<tr>
<td>Apprentice</td>
<td>11 (8.7)</td>
</tr>
<tr>
<td>Employee</td>
<td>58 (46)</td>
</tr>
<tr>
<td>Civil servant</td>
<td>6 (4.8)</td>
</tr>
<tr>
<td>Self-employed</td>
<td>8 (6.3)</td>
</tr>
<tr>
<td>Not employed</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>Retired without income</td>
<td>9 (7.1)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>Not specified</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td><strong>Monthly income (Euro; 1€=US $1.09)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;250</td>
<td>5 (4.0)</td>
</tr>
<tr>
<td>250–499</td>
<td>8 (6.3)</td>
</tr>
<tr>
<td>500–999</td>
<td>17 (13.5)</td>
</tr>
<tr>
<td>1000–1499</td>
<td>12 (9.5)</td>
</tr>
<tr>
<td>1500–1999</td>
<td>13 (10.3)</td>
</tr>
<tr>
<td>2000–2999</td>
<td>26 (20.6)</td>
</tr>
<tr>
<td>3000–3999</td>
<td>16 (12.7)</td>
</tr>
<tr>
<td>4000–4999</td>
<td>6 (4.8)</td>
</tr>
<tr>
<td>&gt;5000</td>
<td>13 (10.3)</td>
</tr>
<tr>
<td>Not specified</td>
<td>10 (7.9)</td>
</tr>
</tbody>
</table>
Ethical Considerations

Our online study was conducted in accordance with the American Psychological Association’s Ethical Principles of Psychologists and Code of Conduct. In particular, data collection was anonymous; harmless to participants; and did not involve deception, injure, or place participants under high levels of physical or emotional stress. In line with 2023 guidelines of the German Research Foundation, formal ethical approval was not required because our study did not include aspects that would necessitate a statement, per subsection two of the “Information on proposals in the field of psychology” [32]. Informed consent was obtained from all participants after the purpose of the study and the data collection were outlined in the survey introduction. Participants indicated their consent by clicking a button. Study data and identifiers were anonymized during the data collection and data analysis to maintain confidentiality. No compensation was awarded to participants.

Results

Table 2 provides an overview of the relative importance values of the attributes and part-worth values of each level for each attribute as determined by conjointly to answer the first research question [29]. A treatment by a physician that is completely compensated by insurance and has no waiting time emerged as the most preferred mode of treatment. Overall, provider was the most important attribute, followed by costs and waiting time. For all attributes, we found two levels with part worths around zero and one positive and negative level. For provider, the physician had a positive part worth and the AI system had a negative part worth, while both the personalized AI and teledermatology had near-zero part worths. For waiting time, immediate results had a positive part worth and a 4-week wait had a negative part worth, while a 1-day and 1-week wait had similar near-zero part worths.

Table 2. Part-worth and relative importance values of the attributes.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Part worth (95% CI)</th>
<th>Relative importance, % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provider</td>
<td></td>
<td>38.6 (35.3 to 41.6)</td>
</tr>
<tr>
<td>AIa</td>
<td>−0.15 (−0.17 to −0.13)</td>
<td></td>
</tr>
<tr>
<td>Personalized AI</td>
<td>−0.06 (−0.08 to −0.04)</td>
<td></td>
</tr>
<tr>
<td>Physician treatment</td>
<td>0.21 (0.19 to 0.24)</td>
<td></td>
</tr>
<tr>
<td>Electronic consultation with physician (teledermatology)</td>
<td>0.005 (−0.01 to 0.02)</td>
<td></td>
</tr>
<tr>
<td>Costs for screening</td>
<td></td>
<td>31.6 (29.0 to 34.0)</td>
</tr>
<tr>
<td>0€ (completely covered by health insurance)</td>
<td>0.15 (0.13 to 0.16)</td>
<td></td>
</tr>
<tr>
<td>10€ copayment</td>
<td>0.06 (0.06 to 0.07)</td>
<td></td>
</tr>
<tr>
<td>25€ copayment</td>
<td>−0.03 (−0.04 to −0.03)</td>
<td></td>
</tr>
<tr>
<td>40€ copayment</td>
<td>−0.18 (−0.19 to −0.16)</td>
<td></td>
</tr>
<tr>
<td>Waiting time for diagnosis</td>
<td></td>
<td>29.8 (27.2 to 32.3)</td>
</tr>
<tr>
<td>Immediate</td>
<td>0.10 (0.09 to 0.11)</td>
<td></td>
</tr>
<tr>
<td>1 day</td>
<td>0.082 (0.07 to 0.09)</td>
<td></td>
</tr>
<tr>
<td>1 week</td>
<td>−0.004 (−0.01 to 0.003)</td>
<td></td>
</tr>
<tr>
<td>4 weeks</td>
<td>−0.18 (−0.20 to −0.17)</td>
<td></td>
</tr>
</tbody>
</table>

aAI: artificial intelligence.
b1€=US $1.09.

Figure 2 shows an overview of the relationships between sociodemographic characteristics and the relative importances to answer the second research question. We found a medium-sized positive relationship between age and provider. In addition, there were two nonsignificant trends. The first indicated an inverse relationship between age and the importance of costs and the second indicated an inverse relationship between income and the importance for costs. All other importances were not systematically related to sociodemographic variables (Table 3).
Figure 2. Relationship between relative importances and sociodemographic characteristics: (A) age, (B) gender, (C) education, (D) income, (E) employment status.
Table 3. Correlation coefficients (Spearman ρ) for the importance values.

<table>
<thead>
<tr>
<th>Sociodemographic characteristics</th>
<th>Relative importance</th>
<th>Provider</th>
<th>Costs for screening</th>
<th>Waiting time for diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Coefficient</td>
<td>0.21&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.17</td>
<td>0.11&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>P value</td>
<td>.02</td>
<td>.05</td>
<td>.25</td>
</tr>
<tr>
<td>Gender</td>
<td>Coefficient</td>
<td>−0.003</td>
<td>−0.03</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>P value</td>
<td>.97</td>
<td>.71</td>
<td>.60</td>
</tr>
<tr>
<td>Education</td>
<td>Coefficient</td>
<td>−0.04</td>
<td>−0.06</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>P value</td>
<td>.64</td>
<td>.45</td>
<td>.45</td>
</tr>
<tr>
<td>Employment status</td>
<td>Coefficient</td>
<td>0.05</td>
<td>−0.11</td>
<td>−0.004</td>
</tr>
<tr>
<td></td>
<td>P value</td>
<td>.53</td>
<td>.22</td>
<td>.96</td>
</tr>
<tr>
<td>Income</td>
<td>Coefficient</td>
<td>0.02</td>
<td>−0.17</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>P value</td>
<td>.81</td>
<td>.07</td>
<td>.24</td>
</tr>
</tbody>
</table>

<sup>a</sup>Pearson correlation coefficient.

Discussion

The aim of this study was to determine how important the attributes provider, costs, and waiting time are for users’ preference for skin cancer screening and to investigate whether sociodemographic characteristics, especially age, are systematically related to participants’ individual importances. We found that provider was as equally important a factor for participants’ decisions as cost and waiting time. While a physician was the most preferred level of this attribute, AI-based treatment was disliked and a personalized AI had the same value for participants as teledermatology. Concerning the relationship between sociodemographic characteristics and relative importances, we found that only age showed a reliable positive association to provider, in which younger participants place less importance on the provider than older participants. In the following, we discuss these findings in turn before discussing the limitations of the study and providing a general outlook.

Regarding the role of the provider in users’ decisions, other studies underline our results that patients exhibit hesitant behavior toward medical AI. Patients would rather not have a treatment than be examined by an AI system, even if the AI system shows the same or better accuracy as a physician [17]. However, the same study also found that patients prefer personalized AI over nonpersonalized AI. Similarly, earlier discrete-choice experiments [6] as well as surveys [33] found that patients prefer a trained medical professional to be involved in their skin cancer screening; 41% of respondents were open to using AI as a standalone system for skin cancer screening and 94% were open to using it as a support system for physicians [33]. Together, existing studies indicate that personalized AI and teledermatology are generally more accepted than nonpersonalized AI for skin cancer screening, while the physician remains the most preferred option.

Concerning the impact of age differences on the acceptance of AI in dermatology, our findings also support some earlier results [21-23]. Higher interest in using teledermatology [21,22] and in using skin cancer–related apps [23] was associated with younger age. The results of cross-sectional studies back up our findings from the choice-based conjoint analysis. Based on these trends, it is possible to imagine that the acceptance of AI in skin cancer screening will rise in the future due to the aging of digital natives and their increased acceptance of AI.

Regarding income and educational factors, our findings do not align with those of previous studies. Ghani et al [22] concluded that higher education attainment and a higher household income increased the interest in using teledermatology. Chang et al [21] came to similar conclusions, indicating that college graduates showed the greatest interest in teledermatology. In addition, Steeb et al [23] showed that a high educational level was associated with a positive attitude toward skin cancer–related apps. While we were not able to show significant relationships to income and educational background, the smaller sample size in this study compared to those of earlier studies might explain this inconsistency.

Previous studies also identified gender differences in the acceptance of AI in skin cancer screening. Chang et al [21] came to the conclusion that men are more likely to use teledermatology than women. Steeb al [23] found similar results in which male gender was significantly associated with a positive attitude toward skin cancer–related apps. However, the gender difference that was reported in earlier studies was not visible in our data. Again, this might be a factor of sample size.
size, but it also might also be that these gender differences reported in earlier questionnaire studies reflect differences in the technology self-concept [34] rather than actual preferences.

Several aspects must be considered in interpretation of our findings. First, the sample was not randomly selected but was based on a convenience sample. While a wide range of recruitment means were used, the results are likely not generalizable to the general public but rather more specific to highly educated young adults. Further research is needed with the target group. Although the sample size may not seem particularly large, sensitivity analysis showed that this sample size was in fact sufficient to detect a medium-sized correlation ($r=0.28$) with a power of 90% and error rate of 5%. Second, some participants contacted us about the meaning of the attribute waiting time because they were unsure whether this pertained to the waiting time for a diagnosis or the waiting time for an appointment. Future studies should make this distinction more explicit to study possible differential effects of these two types of waiting times.

Taken together, we believe that this study adds to the growing body of research using choice-based experiments to investigate the acceptance of AI in health contexts. This approach offers additional insights and is less susceptible to social desirability and other biases [35]. However, the choice-based conjoint analysis only allows studying a small number of potential attributes at a time [24]. Because we included personalized AI as a level for the attribute provider, our study adopted the findings of Longoni et al [17] that personalized AI increases patient acceptance. In addition, we examined factors that may have an impact on patients’ decision-making following the study of Snoswell et al [6].

For the future, it could be interesting to add “AI as a physician support system” to the choice set [33]. It might also be interesting to find out whether patients who perceived themselves as more individualized are less accepting of AI [17]. Additionally, it could be interesting to explore whether specialized knowledge about AI systems would increase patient acceptance [13] and which other factors might have an influence on patients’ acceptance. Ideally, this would not only rely on correlational evidence as used here but also on experimental evidence that shows how preferences and importance may be altered. The variables such as income and educational background cannot be manipulated easily. Nevertheless, we believe that the magnitude of these effects provides some benchmarks for future studies that aim to use experimental methods to alter preferences.

In summary, while there have been technological advances in the effectiveness of AI for supporting skin cancer screening and health care more generally, we believe that the true potential of AI systems can only be realized if patients’ needs and demands are taken into account.

Acknowledgments
This research was conducted within the framework of the project “SAIL: SustAInable Lifecycle of Intelligent Socio-Technical Systems,” funded by the Ministry of Culture and Science of the State of North Rhine-Westphalia under grant NW21-059B. The sole responsibility for the content of this publication lies with the authors. Publication was funded by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) under grant 490988677 and the University of Applied Sciences Bielefeld. Generative artificial intelligence was not used in any portion of the manuscript writing.

Data Availability
The data sets generated and analyzed during this study are available in the zenodo repository [36].

Authors’ Contributions
IJ: study conceptualization, analysis, writing first draft, approval of final manuscript; MS: study conceptualization, analysis, approval of final manuscript; OW: study conceptualization, data collection, approval of final manuscript; GH: study conceptualization, data collection, analysis, approval of final manuscript.

Conflicts of Interest
None declared.

References


28. Orme B. Interpreting the results of conjoint analysis 2010. Sawtooth Software. 2010. URL: https://content.sawtoothsoftware.com/assets/3501e79f-3a9f-4d99-a0f5-1522525e7f3f [accessed 2023-08-08]


31. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]


Abbreviations

AI: artificial intelligence
Digital Health Needs and Preferences During Pregnancy and the Postpartum Period: Mixed Methods Study

Natalie Henrich¹, MPH, PhD; Alison Brinson¹,², MSPH; Alyssa Arnold¹, MPH; Hannah R Jahnke¹, PhD

¹Maven Clinic, New York, NY, United States
²Department of Anthropology, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Abstract

Background: Digital health is increasingly used to meet the needs of perinatal people, with estimates of pregnancy-related internet use ranging from 90% to 97% of pregnant people. As digital health takes on greater importance during the perinatal period, it is essential that providers and developers of digital health content understand why perinatal people use these resources and the features that enhance their experience. However, gaps remain in understanding the content that is most helpful and how the platforms are navigated. Learning directly from perinatal people about their needs will help ensure alignment between perinatal needs and available content.

Objective: This formative study aims to identify the reasons why perinatal people use digital health resources; the features of the digital health platforms that are of greatest importance to them; and how these differ by perinatal stage (pregnancy vs postpartum), mental health conditions, parity, and demographics (race and ethnicity).

Methods: This mixed methods study used interviews; surveys; and secondary data on demographic, health, and pregnancy characteristics to identify the digital health needs and preferences of pregnant and postpartum people who used the Maven digital health platform in the United States during their pregnancy or postpartum period. The interviews informed the content of the surveys and provided additional insights and examples for interpreting the survey results. The surveys were used to collect data from a sample of Maven users, and the results were linked to the secondary data set. The interviews were thematically analyzed, and survey data were analyzed using descriptive statistics and stratified by parity, race, and mental health status.

Results: Overall, 13 people were interviewed (including n=4, 31% pilot interviews), and 147 pregnancy and 110 postpartum survey respondents completed the surveys and had linkable secondary data. Top reasons for using digital health resources during pregnancy were to (1) know what is normal or typical during pregnancy, (2) have access to a health care provider when needed, and (3) know how the baby is developing. Top reasons for postpartum use were to (1) help with breastfeeding, (2) know what normal baby development is, and (3) help with the baby's health issues. Top platform features during pregnancy and the postpartum period were (1) credible and trustworthy information and providers, (2) nonjudgmental information and support, and (3) no cost to the user. In general, more reasons for using digital resources were identified as extremely important during pregnancy compared with postpartum. The results showed minor variations across strata.

Conclusions: This formative research found minor differences in digital resource needs and preferences across user characteristics among perinatal people in the United States. Future work should examine whether there are variations in interests within topics by user characteristics, which may provide additional opportunities to better meet user needs.

doi:10.2196/48960

KEYWORDS
digital health; perinatal; pregnancy; postpartum; interviews; survey; user needs; patient centered; mixed methods
Introduction

Background

Digital health, which refers to digital health applications, ecosystems, and platforms [1-4], is increasingly used to meet the needs of perinatal people [5-8], with estimates of pregnancy-related internet use ranging from 90% to 97% of pregnant people [9-13]. Digital health is expected to have growing importance because of its potential to provide health prevention, consultation, treatment, and management while simultaneously providing an opportunity to reduce costs and improve access to care and patient satisfaction, which together may lead to better informed and more engaged patients [1,4,14,15].

To date, research on perinatal digital health needs falls into 2 categories: self-report via surveys or qualitative data collection (interviews and focus groups) and content analysis of perinatal digital forums. Previous research has shown that access to perinatal digital health information provides reassurance and support [5,16,17]. Both types of studies have found that most people use digital resources during pregnancy to access information on maternal health (eg, pregnancy ailments, pain and complications, and health and nutrition during pregnancy) and during the postpartum period to access information on baby-related topics such as infant sleep and feeding (especially breastfeeding) [18,19]. Self-report studies have also indicated interest in childbirth, fetal and infant development, and infant illnesses, and studies of web-based forums have revealed interest in people and relationships [5,10,11,18-21]. The relative interest in these topics and the specific content within them vary throughout the perinatal period [17,19,20]. Research has also found that a positive experience using digital health resources is characterized by being easy to access and navigate, providing trustworthy and unbiased information, and being customized (eg, trackers and localized information) [4,22,23]. Perinatal digital health users specifically value an experience that provides information that is proactively delivered; offers immediate access; and provides information that is practical and concise, entertaining, and reassuring [5,24,25].

As digital health takes on greater importance during the perinatal period, it is essential that providers and developers of digital health content understand why perinatal people use these resources and what contributes to a positive user experience [26]. The information sought by perinatal people differs from the information that health care providers prioritize and offer. For example, postpartum people seek information on sleep, emotional changes, and breastfeeding, whereas providers focus on medical issues such as bleeding and infection [27,28]. Although there is a growing body of literature, there remain gaps in understanding of the content that is most helpful and how the resources are navigated [5]. There is a demand for more detailed analyses, including which digital resources are most helpful [16], with a particular need to understand what is valued during the postpartum period [18]. Research is needed on digital health needs and experience among perinatal people with mental health conditions, especially pre- and postpartum anxiety [20], and more insights are needed on barriers to and facilitators of using digital health, especially with newer technologies such as apps [4,20]. Learning directly from perinatal people about their needs will help ensure alignment between perinatal needs and available content [18,29]. The design of digital health content should be directly informed by individuals and their experiences. However, the research on patients’ user experience of digital health is limited, especially during the perinatal period [4].

Objectives

This mixed methods study, which used formative interviews and surveys, investigated the reasons why pregnant and postpartum people use digital health resources and what contributes to a positive user experience. Using a national sample of pregnant and postpartum users of Maven [30], a comprehensive reproductive and family digital health platform that includes patient education, care coordination, and provider services, this formative study identified how needs and preferences differ by perinatal stage (ie, pregnancy vs postpartum), mental health conditions, parity, and demographics (race and ethnicity). The results from this study can be used to inform a user-centered approach to developing the content and design of digital health resources and improve the alignment between resources and users. In recognition of the fact that patients have diverse gender identities, we have used gender-inclusive language throughout this paper. If a participant described themselves using gendered language (ie, “as a working mother...”), we used their terminology. When describing the results of previous research, we used the terminology used by the study investigators.

Methods

Study Design

In this mixed methods study, we identified the digital health needs and preferences of pregnant and postpartum people who used the Maven digital health platform during their pregnancy or postpartum period. The study used formative interviews to inform the content of surveys and provide additional insights and examples for interpreting the survey results. Surveys were used to collect data from a sample of Maven users that could be linked to a secondary data set of demographic, pregnancy, and health characteristics in the Maven platform (Figure 1).

https://formative.jmir.org/2024/1/e48960

JMIR Form Res 2024 | vol. 8 | e48960 | p.528
(page number not for citation purposes)
Digital Platform Background
Maven was developed in 2014 as a digital platform to support perinatal people and their families. Users have a primary point of contact called a care advocate, an allied health professional (e.g., social worker or nurse) who works with the user to navigate the services and resources on the platform. Services and resources include web-based classes; educational materials; and internet-based appointments with a diverse team of health care providers, including obstetrician-gynecologists, mental health professionals, midwives, and doulas. Access to Maven is a sponsored benefit through an employer or the health plan of the user or their partner. The study investigators are Maven employees of the Maven clinical research team. When activating their Maven accounts, users consent to the use of their contact information by Maven for various reasons, including research. The investigators only have direct contact with Maven users or access to identifiable data within the scope of specific research projects, in accordance with institutional review board (IRB) approvals, and in concordance with Health Insurance Portability and Accountability Act regulations. Access to user data is minimized by requesting recruitment lists and platform-based data from the Maven product analytics team, which deidentifies the information before sharing it with the research team. To recruit Maven users for interviews and the survey, the recruitment parameters were sent to the product analytics team, which generated a list of deidentified IDs of users who met the parameters. This list of deidentified IDs was sent to the Maven team responsible for direct communication with users, which sent out recruitment emails (using email addresses linked to the deidentified IDs).

Formative Interviews
In September 2022 and October 2022, semistructured interviews with postpartum people were conducted virtually via Zoom (Zoom Video Communications) to identify the reasons why perinatal people use any digital health resources, not just Maven. To help encourage participants to think about digital use broadly, at the start of the interview, they were asked to list all the digital resources used during their pregnancy and the postpartum period, and throughout the interview, they were reminded to think about digital resources in general. The first part of the interview was open-ended, and participants were prompted not only to describe the reasons why they used digital resources but also to identify additional reasons for which they would like to use them. In the second part of the interview, participants were shown (via screen sharing on Zoom videoconferencing) a list of reasons why digital resources could be helpful during the perinatal period. Participants were asked whether they had used a digital resource for each individual reason or whether they would like to use a digital resource for each individual reason in the future. Finally, participants were asked about the importance of digital health resources being inclusive of their entire identity, which was explained as including anything that the participant considered part of their identity, such as race, gender, sexual identity, geography, social or economic status, and mental and physical health conditions, and to describe their experience with (lack of) inclusion.

Pilot interviews were conducted with 4 postpartum Maven staff members who had used Maven during pregnancy and the postpartum period and were post partum at the time of the interview. The interview guide was modified based on their feedback and the interviewer’s learnings from conducting the pilot interviews. Following piloting, interviews were conducted with 9 Maven users, all of whom were post partum so that they could speak to their digital needs and preferences throughout the perinatal period. Participants were recruited with an emphasis on including people from a range of ethnicities and races, health conditions, and parity. Participants were recruited via email. Interviews lasted approximately 1 hour, were conducted via Zoom, and were recorded. An interviewer and a notetaker were present, and transcripts were autogenerated on Zoom. Participants (Maven staff and nonstaff) consented to the interview, had the option of having the video on or off, and could decline to have the interview recorded (one participant...
declined, and we relied on detailed notes for that interview). The consent form was included in the recruitment email and reviewed at the start of the interview, and consent was provided verbally. Participants were given a US $50 Amazon gift card as a thank you for their participation.

The prevalence of each reason for using digital resources was calculated by counting the number of interviews (Maven staff and nonstaff) in which the reason was mentioned as important. The reasons that were mentioned as important in the most interviews were included in the survey.

**Survey Development**

In addition to the semistructured interviews, separate quantitative surveys were developed to inquire about digital resource use during pregnancy and the postpartum period. Through the qualitative analysis, it emerged that participants were describing 2 aspects of digital resource use: the information or support they were seeking and features of the digital health platforms. Consequently, the survey contained 2 sections, each addressing one of these aspects. To focus on common needs and remove outliers that were mentioned by only 1 or 2 participants, reasons for using digital resources and features affecting experience were excluded from the surveys if they were cited by less than one-third of the interview participants. In each section of the survey, participants indicated the importance of each item on a Likert scale that included a description of each option (a little important or doesn’t matter to me—I don’t care much about this, important—I would like to have this but I'm OK without it, very important—I care about this a lot but it’s not absolutely necessary, and extremely important—a “must have”) and subsequently picked the 5 most important items (not ranked). We opted to allow participants to select up to 5 items as a means to balance the tension between the high number of reasons why pregnant and postpartum people are turning to digital resources and the need to prioritize to inform actionable decisions about what content to provide in these resources. To reduce the potential bias of basing responses on their use of Maven, the instructions in the survey reminded respondents to think holistically about digital health resources. The following is an example of the instructions from a section of the pregnancy version of the survey: “Thinking about the reasons you might use any digital resource during pregnancy, rate how much each of these reasons matter to you. Digital resources include websites, apps, videos, blogs, search engines, etc.”

The surveys were piloted with 4 Maven staff members who had used Maven during pregnancy and the postpartum period and were post partum at the time of piloting. The wording and format were modified based on their feedback. The revised surveys were piloted with 4 Maven nonstaff users. During piloting, we identified that nearly all items were selected as very important. To minimize ceiling effects, the response options were revised to provide more response options that reflected higher levels of importance and add definitions (eg, extremely important: a must have). We also opted to reverse the order of the options (with a little important or not important to me as the first option rather than the last). In addition, we included a section for selecting the 5 most important reasons for use and platform features, which would provide information on the relative importance of the items if we continued to have a ceiling effect.

On the basis of feedback from the staff and nonstaff member pilots, we added 2 items that had not emerged in the interviews: using digital resources to help with formula feeding and feeding solid foods. These were added as participants felt that the survey was not inclusive or relevant to the entire postpartum period if it only asked about breastfeeding.

The survey was built in SurveyMonkey (SurveyMonkey Inc) using a Health Insurance Portability and Accountability Act–compliant account.

**Survey Administration**

Survey respondents were recruited from current Maven users. To be eligible for the pregnancy survey, users had to be between 0 and 6 months post partum and have had a Maven account during pregnancy. To be eligible for the postpartum survey, users had to be from 3 to 12 months post partum and have a Maven account during the postpartum period. Recruitment was via email, which included a link to the survey. Each survey link was associated with the user’s unique Maven identifier to link the survey responses to user demographic characteristics previously recorded on the Maven platform (described in the Secondary Data section). A reminder email was sent 1 week after the initial email to anyone who had not clicked through to the survey. The survey remained open for 2 weeks from the time of the initial email. A consent form appeared before the start of the survey, and consent had to be provided digitally to access the survey. Participants were entered into a draw for a chance to win 1 of 5 US $100 Amazon gift cards per survey (5 for the pregnancy survey and 5 for the postpartum survey). Data were collected from January 18 to 29, 2023.

**Secondary Data**

Maven users are encouraged but not required to fill out a questionnaire when they activate their account. The Maven pregnancy questionnaire collects data on physical and mental health, pregnancy conditions, demographics, and parity. The Maven postpartum questionnaire collects information on demographics, mode of delivery, and previous pregnancy conditions. For survey participants who completed the Maven questionnaires, we linked their survey responses with their Maven questionnaire data using their unique Maven identifier. Maven users consent to the use of their Maven questionnaires for research when they activate their accounts, and the research team does not have access to any identifying information.

**Analyses**

**Survey Analysis**

Respondents who completed the survey and a Maven questionnaire were included in the survey analysis. Figure 2 shows the pathway for inclusion in the analysis, from eligibility to be recruited to starting and completing the survey and completing a Maven questionnaire. The quantitative data from both the pregnancy and postpartum surveys were assessed primarily through the evaluation of descriptive statistics. Chi-square and Fisher exact tests were used to assess differences among pregnancy items, postpartum items, and user experience...
features by parity, presence of any mental health conditions, and race and ethnicity. A user was considered to have a mental health condition if they reported any of the following: a history of anxiety, depression, perinatal mood disorder, or high pregnancy-related anxiety. A history of anxiety or depression was assessed in the Maven questionnaires using the following question—“Do any of these conditions apply to you or did they in the past?”—with the selection of Anxiety or depression from a list of conditions. Experience of perinatal mood disorder was assessed using the following question—“Have you experienced any of the following during this pregnancy?”—with a selection of Perinatal mood disorder. Pregnancy-related anxiety was assessed on a 5-item Likert scale in response to the following question—“On a scale of 1-5, how anxious are you feeling about your pregnancy?”—with responses of 4 (very) or 5 (extremely) indicating the presence of pregnancy-related anxiety. For analyses stratified by race and ethnicity, we recategorized individuals who identified as Hispanic, Black, Asian or Pacific Islander, or multiracial into 1 category labeled non-White because of our small sample of people of racial and ethnic minorities. Consistent with the study’s aim of identifying differences in needs and preferences between respondents with these characteristics, P values were calculated to determine whether the differences were statistically significant (with results considered to be statistically significant if the P value was <.05). Statistical analyses were conducted using RStudio (Posit Software, PBC).

**Figure 2.** Flowchart of the pregnancy survey respondents (A) and postpartum survey respondents (B) included in the survey analysis.

### Interview Analysis

The interviews were analyzed thematically using a framework analysis [31-36]. Themes were identified deductively based on the interview guide and inductively based on the messages that emerged from the interviews. Themes and their definitions were reviewed by all team members who participated in the interviews (as interviewers or notetakers) and were modified based on group discussion and consensus. The analysis table was populated based on the notes taken during the interviews and subsequent analysis discussions. Audio recordings and transcripts were referenced if the notes were unclear or lacked sufficient details to populate the analysis table. All entries in the analysis table were reviewed by NH, who attended all but one of the interviews. Any information that was missing or potentially miscategorized in the table was flagged for review by the person who entered the information and, if necessary, discussed to reach a consensus. To better understand the experiences and perspectives associated with the topic of each of the items in the quantitative surveys, the corresponding interview themes were summarized, and representative anecdotes or quotes were selected. All quotes were verified using the audio recordings. Quotes and anecdotes are cited using the interview participant number (eg, P1 refers to participant 1), and it is specified whether the participant was from the pilot interviews (eg, Pilot P1 refers to the first interview among the pilot participants).

### Ethics Approval

The study protocol was approved by the WCG IRB (study 1338443). The use of secondary data was deemed exempt by the WCG IRB under 45 Code of Federal Regulations § 46.104(d)(4).

### Results

#### Survey Sample Characteristics

In total, we had 147 pregnancy survey respondents and 110 postpartum survey respondents with complete Maven questionnaires (Table 1 and Figure 2). All survey responses
were collected within 48 hours of sending the initial recruitment email. The average age of the respondents was 32.8 (SD 4.6) years and 32.7 (SD 3.5) years for pregnancy and postpartum survey respondents, respectively. On the basis of those who reported their race and ethnicity, our sample was comparable with the overall US population in terms of the percentage of people identifying as non-Hispanic White (134/257, 52.1%) and multiracial (5/257, 1.9%), with an overrepresentation of people identifying as Asian (31/257, 12.1%) and an underrepresentation of people identifying as Black (4/257, 1.6%) or Hispanic (16/257, 6.2%) [37]. A total of 44.9% (66/147; pregnancy survey) and 51.8% (57/110; postpartum survey) of the respondents had household incomes of >US $100,000. Approximately 28% (72/257) of the sample reported a history of a mental health condition (Table 1).

Table 1. Characteristics of respondents who took the pregnancy survey and the postpartum survey.

<table>
<thead>
<tr>
<th>Demographic characteristics</th>
<th>Pregnancy survey respondents (n=147)</th>
<th>Postpartum survey respondents (n=110)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y), mean (SD)</td>
<td>32.8 (4.6)</td>
<td>32.7 (3.5)</td>
</tr>
<tr>
<td><strong>Race and ethnicity, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>8 (5.4)</td>
<td>8 (7.3)</td>
</tr>
<tr>
<td>Non-Hispanic Asian</td>
<td>21 (14.3)</td>
<td>10 (9.1)</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>3 (2)</td>
<td>1 (0.9)</td>
</tr>
<tr>
<td>Non-Hispanic multiracial or American Indian</td>
<td>3 (2)</td>
<td>2 (1.8)</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>77 (52.4)</td>
<td>57 (51.8)</td>
</tr>
<tr>
<td>I prefer not to say</td>
<td>35 (23.8)</td>
<td>17 (15.5)</td>
</tr>
<tr>
<td><strong>BMI (kg/m²), n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤18.5</td>
<td>4 (2.7)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>18.5 to 24.9</td>
<td>76 (51.7)</td>
<td>49 (44.5)</td>
</tr>
<tr>
<td>25 to 29.9</td>
<td>30 (20.4)</td>
<td>15 (13.6)</td>
</tr>
<tr>
<td>≥30</td>
<td>29 (19.7)</td>
<td>15 (13.6)</td>
</tr>
<tr>
<td>Social Vulnerability Index (high), n (%)</td>
<td>10 (6.8)</td>
<td>2 (1.8)</td>
</tr>
<tr>
<td><strong>Household income (US $), n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20,000-34,999</td>
<td>1 (0.7)</td>
<td>1 (0.9)</td>
</tr>
<tr>
<td>35,000-49,999</td>
<td>3 (2)</td>
<td>2 (1.8)</td>
</tr>
<tr>
<td>50,000-74,999</td>
<td>7 (4.8)</td>
<td>3 (2.7)</td>
</tr>
<tr>
<td>75,000-99,999</td>
<td>18 (12.2)</td>
<td>9 (8.2)</td>
</tr>
<tr>
<td>≥100,000</td>
<td>66 (44.9)</td>
<td>57 (51.8)</td>
</tr>
<tr>
<td>I prefer not to say</td>
<td>27 (18.4)</td>
<td>10 (9.1)</td>
</tr>
<tr>
<td>Nulliparous (yes), n (%)</td>
<td>97 (66)</td>
<td>58 (80.6)</td>
</tr>
<tr>
<td><strong>Mental health history, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>33 (22.4)</td>
<td>18 (25)</td>
</tr>
<tr>
<td>Depression</td>
<td>14 (9.5)</td>
<td>4 (5.6)</td>
</tr>
<tr>
<td>Perinatal mood disorder</td>
<td>1 (0.7)</td>
<td>2 (2.8)</td>
</tr>
</tbody>
</table>

*aVariables with <15% of missing data: age, race, social vulnerability, BMI (pregnancy survey respondents), mental health history (pregnancy survey respondents), and parity (pregnancy survey respondents). Variables with 15% to 30% of missing data: household income and BMI (postpartum survey respondents). Variables with 30% to 40% of missing data: parity (postpartum survey respondents) and mental health history (postpartum survey respondents).*

*bMeasured using the Centers for Disease Control and Prevention Social Vulnerability Index.*

*cA total of 65.5% (72/110) of postpartum survey respondents provided parity or mental health data. Therefore, the sample size for those variables is 72.*

**Reasons for Use of Digital Resources During Pregnancy**

On the basis of the survey responses, the most important reasons for using digital resources during pregnancy were to know what is normal or typical during pregnancy, have access to a health care provider when needed, and know how the baby is changing and developing during pregnancy (Table 2).
Table 2. Reasons identified as extremely important when using digital health resources during pregnancy by parity, mental health status, and race and ethnicity (N=147).

<table>
<thead>
<tr>
<th>Reasons for using digital health resources during pregnancy</th>
<th>Overall, n (%)</th>
<th>Parity</th>
<th>Mental health</th>
<th>Race and ethnicitya</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Parous (n=50), n (%)</td>
<td>Nulliparous (n=97), n (%)</td>
<td>No mental health history (n=105), n (%)</td>
</tr>
<tr>
<td>To know what is normal or typical during pregnancy</td>
<td>95 (64.6)</td>
<td>24 (48)</td>
<td>71 (73.2)</td>
<td>68 (64.8)</td>
</tr>
<tr>
<td>To have access to a health care provider when I need them</td>
<td>84 (57.1)</td>
<td>26 (52)</td>
<td>58 (59.8)</td>
<td>57 (54.3)</td>
</tr>
<tr>
<td>To know how my baby is changing and developing during pregnancy</td>
<td>76 (51.7)</td>
<td>20 (40)</td>
<td>56 (57.7)</td>
<td>58 (55.2)</td>
</tr>
<tr>
<td>To help me manage discomfort during pregnancy</td>
<td>61 (41.5)</td>
<td>19 (38)</td>
<td>42 (43.3)</td>
<td>44 (41.9)</td>
</tr>
<tr>
<td>To become more empowered to speak up for myself during pregnancy and labor</td>
<td>56 (38.1)</td>
<td>16 (32)</td>
<td>40 (41.2)</td>
<td>41 (39)</td>
</tr>
<tr>
<td>To explore my options for labor and delivery</td>
<td>55 (37.4)</td>
<td>14 (28)</td>
<td>41 (42.3)</td>
<td>42 (40)</td>
</tr>
<tr>
<td>To help me manage anxiety or depression</td>
<td>55 (37.4)</td>
<td>19 (38)</td>
<td>36 (37.1)</td>
<td>37 (35.2)</td>
</tr>
<tr>
<td>To help me manage my health problems (chronic, ongoing, or related to pregnancy)</td>
<td>45 (30.6)</td>
<td>17 (34)</td>
<td>28 (28.9)</td>
<td>32 (30.5)</td>
</tr>
<tr>
<td>To get additional information or clarification on things my doctor tells me</td>
<td>44 (29.9)</td>
<td>15 (30)</td>
<td>29 (29.9)</td>
<td>32 (30.5)</td>
</tr>
<tr>
<td>To know the size of my baby throughout pregnancy</td>
<td>35 (23.8)</td>
<td>14 (28)</td>
<td>21 (21.6)</td>
<td>29 (27.6)</td>
</tr>
<tr>
<td>To help me with healthy eating during pregnancy</td>
<td>32 (21.8)</td>
<td>7 (14)</td>
<td>25 (25.8)</td>
<td>27 (25.7)</td>
</tr>
<tr>
<td>To help me with being physically active during pregnancy</td>
<td>30 (20.4)</td>
<td>9 (18)</td>
<td>21 (21.6)</td>
<td>26 (24.8)</td>
</tr>
<tr>
<td>To get recommendations of things to buy to care for my baby or myself</td>
<td>25 (17)</td>
<td>7 (14)</td>
<td>18 (18.6)</td>
<td>20 (19)</td>
</tr>
<tr>
<td>To connect with a community of people who are in the same stage of pregnancy as me</td>
<td>17 (11.6)</td>
<td>6 (12)</td>
<td>11 (11.3)</td>
<td>11 (10.5)</td>
</tr>
</tbody>
</table>

*Respondents who selected I prefer not to say for race and ethnicity were not included in this comparison.

In interviews, participants described the role that digital resources play in meeting these needs (Table 3), and they emphasized the importance of these resources during their first pregnancy (Pilot P1 and P2). Among the information they sought about what is normal, participants wanted to know what they would encounter during each phase of labor and delivery and were looking for information such as how long to expect to push. A participant described learning about the phases during a digital birth-planning appointment:

[The digital appointment with the birth planning specialist] was very helpful to know what to expect at the hospital during birth, and actually I feel like they were more helpful than the in person doula that I hired...I remember them stepping through the birth process. You know, active labor and transition, the pushing and talking about the ring of fire. [P5]
Table 3. Illustrative quotes for using digital resources (pregnancy and postpartum interviews).

<table>
<thead>
<tr>
<th>Reasons for using digital health resources</th>
<th>Illustrative quote</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>During pregnancy</strong></td>
<td></td>
</tr>
<tr>
<td>To know the size of my baby throughout pregnancy</td>
<td>“What is the size of my baby today? Let’s see. I will also open the app and see. So that was a good thing as well. So this is like ‘What vegetable is your baby right now?’” (P2)</td>
</tr>
<tr>
<td>To help me manage discomfort during pregnancy or postpartum</td>
<td>“[The digital resource helped with] cracked and bleeding nipples” (P9).</td>
</tr>
<tr>
<td>To become more empowered to speak up for myself during pregnancy and labor</td>
<td>“I read a ton of articles [online] because I felt like otherwise it was like the blind leading the blind. My husband had never done that before. I’m an only child like my mom was going to be there, but she hadn’t done it in thirty years. So it was kind of like, Okay, Who in this room knows what they’re doing? The answer was going to be nobody...So I read a lot just to know these are all the possible outcomes. This is how it might go. These are the things that I’m going to be asked, and I wanted to know, while I’m in a calm and sound state of mind, like is my answer Yes? Is my answer No? What are the implications if I say yes or no. And then I [can] communicate that to my partner and my mother. So if they needed to advocate for me [they could]” (P4).</td>
</tr>
<tr>
<td>To explore my options for labor and delivery</td>
<td>“[Exploring my labor and delivery options] I would say [getting support on Maven] helped a little but that was more so because of me, and I didn’t look into it as much as I should have...Because, I thought I would have a natural birth. But I ended up having a C section, but just not knowing what to expect, and more so it was on the process of like what an induction is, and like how your body would look like afterwards, because I didn’t get to in my entire pregnancy. But all the drugs they pumped into you to induce you caused me to blow up after, and I just didn’t expect that, no one tells you, right? They just tell you to consent, and you say yes, so like all that, I wish I knew. But I don’t think that I looked into it as much, because I just didn’t know” (P8).</td>
</tr>
<tr>
<td><strong>During the postpartum period</strong></td>
<td></td>
</tr>
<tr>
<td>To help me with feeding my baby breast milk</td>
<td>“I didn’t see lactation consulting in person until after birth, which was at the pediatricians office. So it’s nice to have that, to be able to talk to one [online] beforehand, to know what to prepare for. And then also I talked to one after as well, after giving birth. That was super helpful” (P5).</td>
</tr>
<tr>
<td>To help me with my baby’s health issues</td>
<td>“[My baby] was very colicky, I took an [online] appointment and showed the doctor what was happening, so they also gave me a hint she might have a milk allergy, [and said to] check with your doctor and check the stools and all” (P2).</td>
</tr>
<tr>
<td><strong>During pregnancy or the postpartum period</strong></td>
<td></td>
</tr>
<tr>
<td>To know what is normal or typical during pregnancy or postpartum</td>
<td>“[Getting digital help on my baby’s sleep] helped a lot, because I had a lot of questions about what’s normal, what’s not normal and they answered everything. It made me feel like I wasn’t doing something bad” (P8).</td>
</tr>
<tr>
<td>To know how my baby is changing and developing during or after pregnancy</td>
<td>“I had like three apps on my phone to track [how my baby was changing] like every week” (P4).</td>
</tr>
<tr>
<td>To have access to a health care provider when I need them</td>
<td>“It wasn’t just 8 to 5 that I could access someone...The sleep consultant was like you can just message me any time, and I’ll respond and help you real time with what you need. That’s something that you can’t get with any in person care provider” (P3).</td>
</tr>
<tr>
<td>To help me with being physically active during pregnancy or the postpartum period</td>
<td>“I was trying to be physically active during pregnancy but then there were times where I stopped because of the spotting. They didn’t know what was going on. Then I found out it was okay. But then the last trimester it got tougher since I was tired a lot and having pain...It would have been nice to see what I could do, like modifications and exercises I could do around the pain I was having” (P5).</td>
</tr>
<tr>
<td>To help me with healthy eating during pregnancy or the postpartum period</td>
<td>“Just offering ideas of foods, more nutritious foods to eat, healthier options, because you can look around on the Internet but just talking to a person who is, you know, that’s their specialty, and you can give them personal information about yourself so they can help you” (P9).</td>
</tr>
<tr>
<td>To help me manage anxiety or depression</td>
<td>“I was in so much pain and I thought maybe [the baby will] be also in pain, or somehow it will affect the baby, because I’m not happy. So then I talked to [an online] doctor, and one of the doctors was very good and she talked to me more than the time of appointment, and she was very kind, and she said, I know how you are feeling, but she gave me so much support, emotional support, and she said, ‘You can come to talk to me anytime’. ...So she then she assured me, ‘Everything is fine, and you don’t need to worry’” (P2).</td>
</tr>
<tr>
<td>To help me manage my health problems (chronic, ongoing, or related to pregnancy or postpartum)</td>
<td>“I have arthritis and Hashimoto. So two autoimmune diseases. So when I got pregnant, I had a lot of concerns about that, and how that would affect my pregnancy, or vice versa” (P8).</td>
</tr>
<tr>
<td>To get recommendations of things to buy to care for my baby or myself</td>
<td>“I’m a working mother. At the time, you know, working while pregnant, and so I didn’t have time to spend hours and hours going down the rabbit hole of like, this stroller versus that stroller. Like just like, tell me what you recommend, and we’re moving on. So it felt like the research had been done for you” (P4).</td>
</tr>
</tbody>
</table>
In interviews, participants talked about being able "to count on" accessing health care providers on the internet for urgent situations, such as a participant (P1) who used a digital health care provider when she did not feel her baby kick for a while and wanted to know whether she should go to the emergency room.

Almost all the participants talked about the importance of knowing how their baby was growing and developing during pregnancy. They described wanting to know about the development week by week and found that digital resources could easily provide this information.

In the survey, nulliparous respondents were more likely than parous respondents to find knowing what was normal (P=0.004) and how their baby was developing (P=0.06) to be extremely important (Table 2). Compared with non-Hispanic White respondents, respondents identifying as Hispanic or people of racial and ethnic minorities were more likely to report that digital health resources were extremely important for the management of anxiety or depression (P=0.003) and chronic health problems (P=0.002). Compared with respondents without a mental health history, respondents with a mental health history were more likely to find it important to receive web-based help with healthy eating during pregnancy (P=0.04; Table 2).

**Reasons for Use of Digital Resources During the Postpartum Period**

On the basis of survey responses, during the postpartum period, the most important reasons for using digital resources were to help with feeding the baby breast milk, know what is normal baby development (eg, developmental milestones), help with the baby’s health issues, and help with feeding the baby solid foods (Table 4).

The reasons for using digital resources during the postpartum period were described in interviews (Table 3). When asked specifically about the use of digital health resources for help with feeding, interview participants primarily spoke about support for breastfeeding. In particular, participants found it valuable to connect with a lactation consultant during pregnancy to prepare for breastfeeding and during the postpartum period for tips that facilitated feeding:

"I didn’t see lactation consulting in person until after birth, which was at the pediatrician’s office. So it’s nice to have that, to be able to talk to one [online] beforehand, to know what to prepare for. And then also I talked to one after as well, after giving birth. That was super helpful. Just things I could do for milk supply. And then the one major helpful thing that I learned is about the pump parts where I can just store them in the fridge, and I don’t have to wash it after every time, because it was a hassle, too." [P5]

In interviews, participants described using digital resources to know what to expect about life with a baby and to find out whether the things they encountered were normal, such as struggling with the baby’s sleep schedule, and what it is really like to breastfeed. Participants voiced frustration that this type of information was not provided by in-person providers, so they looked for answers on the internet:

"They don’t tell you anything [at the hospital]. A big one for me was breastfeeding, like just how complex it is. There’s just so many things in the hospital they don’t tell you anything, either. They just ask you like, “Are you going to breastfeed, yes or no,” and you have to decide on the spot, and, like you don’t know anything about this. After that, coming home, I definitely did research on [a digital platform] and the resources that they had for help." [P6]

Participants also wanted digital resources to provide information on what is normal for babies as they grow, including what developmental milestones to expect at each age. A participant described using an app to know what to expect as her baby changed:

"[There is an app that when you put in the baby’s due date, this helps you to see your baby’s development, like milestone-wise. So week by week, you know some of the things to look for. And mainly I use it to see like, when are one of the cranky times coming up because they’re going through a milestone leap where they’re learning, you know lots of new things." [P9]

In interviews, participants provided anecdotes about the value of digital resources for helping with the baby’s health issues, especially the ability to access support when it is needed. A new parent described meeting with a digital sleep coach in the middle of the night when her baby would not sleep, a service she said she could not obtain from her in-person providers as they only handled emergencies after hours:

"It wasn’t just 8 to 5 that I could access someone...The sleep consultant was like you can just message me any time, and I’ll respond and help you real time with..."
what you need. That’s something that you can’t get with any in person care provider. If you call their after hours number it’s because it’s an emergency right? Not because you are just, like, wondering why your baby is fussy, or awake window, or, you know whatever. [P3]

In the survey, nulliparous respondents were more likely than parous respondents to find it extremely important that digital resources help with feeding their baby solid food ($P=0.01$) and with their own health problems ($P=0.06$; Table 4). Respondents identifying as Hispanic or people of racial and ethnic minorities were more likely than non-Hispanic White respondents to report that it is extremely important that digital resources can be used to connect them with a community of people who have a baby of the same age ($P=0.02$) and to obtain recommendations of things to buy to care for their baby or themselves ($P=0.08$). Respondents with a mental health history were more likely than those without a mental health history to find it important to obtain web-based help with breastfeeding their baby ($P=0.01$; Table 4).

Table 4. Reasons identified as extremely important when using digital health resources during the postpartum period by parity, mental health status, and race and ethnicity (N=110)$^b$.

<table>
<thead>
<tr>
<th>Reasons for using digital health resources during the postpartum period</th>
<th>Overall, n (%)</th>
<th>Parity</th>
<th>Mental health</th>
<th>Race and ethnicity$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nulliparous (n=58), n (%)</td>
<td>non-Hispanic White (n=57), n (%)</td>
<td>No mental health history (n=49), n (%)</td>
<td>People of racial and ethnic minorities (n=21), n (%)</td>
</tr>
<tr>
<td>To help me with feeding my baby breast milk</td>
<td>64 (58.2)</td>
<td>6 (42.9)</td>
<td>37 (63.8)</td>
<td>0.12</td>
</tr>
<tr>
<td>To know what is normal baby development (milestones)</td>
<td>54 (49.1)</td>
<td>7 (50)</td>
<td>30 (51.7)</td>
<td>0.89</td>
</tr>
<tr>
<td>To help me with my baby’s health issues</td>
<td>54 (49.1)</td>
<td>6 (42.9)</td>
<td>32 (55.2)</td>
<td>0.38</td>
</tr>
<tr>
<td>To help me with feeding my baby solid foods</td>
<td>53 (48.2)</td>
<td>4 (28.6)</td>
<td>36 (62.1)</td>
<td>0.01</td>
</tr>
<tr>
<td>To help me manage anxiety or depression</td>
<td>45 (40.9)</td>
<td>5 (35.7)</td>
<td>22 (37.9)</td>
<td>0.77</td>
</tr>
<tr>
<td>To help me with feeding my baby formula</td>
<td>30 (27.3)</td>
<td>2 (14.3)</td>
<td>18 (31)</td>
<td>0.20</td>
</tr>
<tr>
<td>To get additional information or clarification on things my physician or pediatrician tells me</td>
<td>29 (26.4)</td>
<td>4 (28.6)</td>
<td>14 (24.1)</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>To help me with my health problems (chronic, ongoing, or related to pregnancy or postpartum)</td>
<td>22 (20)</td>
<td>0 (0)</td>
<td>12 (20.7)</td>
<td>0.06</td>
</tr>
<tr>
<td>To help me with being physically active after my baby is born (postpartum exercise)</td>
<td>18 (16.4)</td>
<td>4 (28.6)</td>
<td>10 (17.2)</td>
<td>0.47</td>
</tr>
<tr>
<td>To help me with healthy eating after my baby is born</td>
<td>16 (14.5)</td>
<td>2 (14.3)</td>
<td>10 (17.2)</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>To help me with losing weight after my baby is born</td>
<td>12 (10.9)</td>
<td>1 (7.1)</td>
<td>7 (12.1)</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>To obtain recommendations of things to buy to care for my baby or myself</td>
<td>11 (10)</td>
<td>1 (7.1)</td>
<td>9 (15.5)</td>
<td>0.67</td>
</tr>
<tr>
<td>To connect with a community of people who have a baby the same age as mine</td>
<td>9 (8.2)</td>
<td>0 (0)</td>
<td>6 (10.3)</td>
<td>0.33</td>
</tr>
</tbody>
</table>

$^a$A total of 65.5% (72/110) of postpartum survey respondents provided parity or mental health data.

$^b$Respondents who selected “I prefer not to say” for race and ethnicity were not included in this comparison.
The reasons for using digital resources that the fewest survey respondents selected as extremely important during the perinatal period were uses related to general wellness (ie, healthy eating, exercise, and losing weight), connecting with a community in the same perinatal stage, and obtaining recommendations of things to buy to care for their baby or themselves. Notably, these lowest-ranked reasons were still selected by 8.2% (9/110) to 22% (32/147) of the respondents (Tables 2 and 4).

In contrast to the surveys, help with healthy eating was one of the topics that the most interview participants identified as a reason for using digital resources, although it was usually in response to specifically being asked about it in the structured part of the interview. When discussing the use of digital resources for healthy eating, participants tended to refer to access to personalized information such as food recommendations that fit with their dietary restrictions. A participant described why tailored information would be helpful to her on a digital health platform:

*Because if I had dietary restrictions I’m going to have to go searching more for that information on my own.*

*P9*

### Importance of Digital Health Platform Features During Pregnancy and the Postpartum Period

The 3 most important features of digital health platforms identified through both the pregnancy and postpartum surveys were credible and trustworthy information and providers, digital resources that are free to the user, and nonjudgmental information and support. Although access to appointments (ie, fast access and access at convenient times) was not among the top 3 most important features, approximately half of the respondents felt that it was extremely important (Table 5). See Multimedia Appendices 1 and 2 for digital health platform features identified as extremely important by parity, mental health status, and race and ethnicity during pregnancy and the postpartum period, respectively.

<table>
<thead>
<tr>
<th>Table 5. Digital health user experience features identified as extremely important during pregnancy and the postpartum period.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital health platform features</td>
</tr>
<tr>
<td>Credible and trustworthy information and providers</td>
</tr>
<tr>
<td>Nonjudgmental information and support</td>
</tr>
<tr>
<td>Digital resources that are free to me</td>
</tr>
<tr>
<td>Fast access to appointments</td>
</tr>
<tr>
<td>Easy to find information; easy to navigate</td>
</tr>
<tr>
<td>Access to appointments at convenient times</td>
</tr>
<tr>
<td>Information that is actionable (specific recommendations for what to do)</td>
</tr>
<tr>
<td>Receive fast responses to my digital messages</td>
</tr>
<tr>
<td>Resources that are specific to my needs (personalized)</td>
</tr>
<tr>
<td>Access to a lot of information on each topic (depth of information)</td>
</tr>
<tr>
<td>Access to information on a lot of topics (breadth of topics)</td>
</tr>
<tr>
<td>Consistent care or support from the same people over time on the digital platform</td>
</tr>
<tr>
<td>Proactive outreach from the digital resource (pushes content to me; provider reaches out to me)</td>
</tr>
<tr>
<td>Care or content that fits with my culture and identity</td>
</tr>
</tbody>
</table>

Interview participants described why the features of digital health resources were important (Textbox 1). When looking to interact with content that was trustworthy, participants mentioned a range of ways in which they determined whether digital content was credible. A couple of participants explained that they trusted the content if it was consistent with the information they received from their in-person providers. This was done by taking the information from the digital content to the in-person provider for verification or by seeing that the content was consistent with what they had already been told:

*Whatever I heard from the [web-based] doctor, it was matching with my experience and my other doctor as well...it builds trust since my doctor is saying the same thing.* [P2]

*So it was like things that the [web-based] practitioner might have suggested or recommended that I would then mention at my doctor’s appointment...kind of to get a check box from my regular doctor, but it was always like, she agreed with what had been communicated.* [P4]
Textbox 1. Illustrative quotes regarding features of digital health resources (pregnancy and postpartum interviews).

**Credible and trustworthy information and providers**
- “Sleep schedules, again, with so much information online, you get really caught up in, like, ‘Am I doing the right thing?’ especially in the beginning. I remember one thing, I was struggling with my baby’s sleep schedule. And so I went to the sleep expert from [my digital platform]. And they said that’s normal, like, a baby shouldn’t really have a sleep schedule until six months. They’ll never have one, and it’s like everything online, says the opposite! So just like having information that is accurate and like coming from experts [really helped]” (P8).

**Digital resources that are free to me**
- “But because this service was free to me I felt like ‘well, it’s there, and I might as well use it,’ so I’ve taken advantage of it” (P3).

**Nonjudgmental information and support**
- “And these platforms really help because, one, you’re able to speak to someone different that may not even have your medical history. They don’t have your medical history in front of you. So maybe it’s the medical system, health care system thinking or judging you based on your history, right, rather than just listening to this specific one time incident. So maybe they’re less biased ‘cause they just have the one thing they’re looking at” (P8).

**Fast access to appointments**
- “If I’m struggling with a situation that I think is super important with the baby like the baby has a cold and has a runny nose, fever, I’m gonna want whoever is available next” (P9).

**Easy to find information; easy to navigate**
- “[My app] did not help [for finding information on baby’s development] because the information, I guess I didn’t seek it out, but I also didn’t see it readily available” (P9).

**Access to appointments at convenient times**
- “Especially for a new mom, if you’re trying to work around is my baby napping or not, and that’s not always predictable, just to be able to quickly touch base with someone. I used it during pregnancy as well, so I feel it kind of rounds out the care you are receiving from your in person doctor” (P4).

**Information that is actionable (specific recommendations for what to do)**
- “In my state it is legal to end the pregnancy, right? So [the online provider] gave me the support that if you want to go for this, this will happen, or if you go, these are the steps you need to follow, and all...She was telling me all the steps, what I need to do, and what I not to do, and she gave me a support for keeping this pregnancy” (P2).

**Receive fast responses to my digital messages**
- “When I messaged, it wasn’t always [the same person] that responded, it was someone else. But it’s nice to have that, too. Like, someone quickly getting back to me. So I didn’t mind that” (P5).

**Resources that are specific to my needs (personalized)**
- “Because if I had dietary restrictions I’m going to have to go searching more for that information on my own. [And it would be helpful to] come up with a list of foods that would be specific to me, because we didn’t do that. I realized that may be asking a lot of someone. But that’s just what I see would be super helpful” (P9).

**Access to a lot of information on each topic (depth of information)**
- “I like the options when I went in, and I would read through an article. It’s like, Oh, here’s some other related stuff to this. If you’re still interested, I really like that because there was a couple of times with a few articles where, like I want to know more about this, and so I would keep clicking and read a little bit more” (P6).

**Access to information on a lot of topics (breadth of topics)**
- “[It would help if the platform had], like, a more robust portfolio of different like classes and articles” (P4).

**Consistent care or support from the same people over time on the digital platform**
- “The sleep coach that I am talking to now. She’s the same one I’ve talked to every single time...[It’s important to have continuity], especially just like to not have to answer the same questions over and over again, like they just kind of know who I am, looking for me, my style, and just more get to the care rather than have to do the basic information each session” (P3).

**Proactive outreach from the digital resource (pushes content to me; provider reaches out to me)**
- “One thing I like about the Baby Center or a lot of the articles I get from other providers is that they’re just in the email, and I don’t have to go anywhere” (P7).
Care or content that fits with my culture and identity

- “The [online doctor] knew Indian culture. So they knew Indian body like how our body reacts after [giving birth]. We are different than other people. So by giving specific things or a suggestion to Indians that is specific to my body, like after delivery, Indians don’t drink too much water. They understand, but my [in-person] doctor will say, drink four litre five litre today. I know so but one of the doctor on Maven when I talk, she said, I know the Indian people don’t drink too much water just after delivery, so I understand that. So if you want to have less water, so you can have a coconut water instead of normal water that will increase the hydration. So that kind of suggestion I don’t get from normal doctor” (P2).

Others said that they trusted digital content that was reviewed by medical professionals (Pilot P3) or had a “provider stamp of approval” (P9). A couple of participants were looking for digital resources that provided information from the people who conducted the research (P8) and that included “the latest research and peer reviewed evaluation” (P3).

In interviews, participants explained that access to free digital health services enables users to consume more health resources and spend more time interacting with health care providers than during in-person appointments:

[When there are] no limit on appointments...suppose I forgot something and I want another appointment and I can do it today. I don’t have to wait for another appointment. [P2]

They also accessed types of resources that they valued but would not use if there was an out-of-pocket cost:

So I think I just wouldn’t have gotten what I needed [in-person] because I have two other children, and I didn’t [get what I needed] with them. Just, like you know, Google things or Pinterest, or talk to friends, and that’s it. But because this service was free to me I felt like “well, it’s there, and I might as well use it,” so I’ve taken advantage of it. And so things like a sleep consultant I probably wouldn’t have paid for. But it was just so nice to have a sounding board...So yeah, I think just having services there that I wouldn’t have made the effort to use before, but because they’re just sitting there on my phone, I might as well, and it’s free to me. [P3]

Participants highlighted that free digital resources are particularly important as the cost of in-person services is very high and they were looking for a cheaper alternative:

Free resources—that’s big, since in person visits are very expensive. [P5]

Interview participants also wanted to interact with digital health content and people who provided information and support without judgment. A participant explained that it is easier to receive nonjudgmental care from a digital health care provider as the person only knows the information that is being shared in the moment and cannot be biased by previous interactions or conditions. Another participant explained that she did not want to discuss her baby’s sleep issues with her in-person pediatrician as she was worried that the physician would judge her as a bad mother. It felt safer to explore the problem with the web-based provider, with whom she did not have an ongoing relationship (Pilot P4).

Interview participants shared anecdotes illustrating the accessibility of web-based providers and described connecting with these providers at night and on weekends or around their work or family schedules:

Whereas in person, they’re always busy. So there’s some waiting time...If I had concerns, I want to consult with somebody...[it’s] very helpful where I can just find someone to talk to within the week, sometimes even the same day...Being able to get in touch with someone quickly during pregnancy and even postpartum has been super helpful. [P5]

There was one time where I had a concern and it was 2AM and I reached out to a practitioner over the phone, it might have been with Maven or it might have been with Telehealth, and [I had] flexibility with getting help at abnormal times of day. [P8]

In general, the importance of digital health platform features was consistent by race, parity, and mental health. Differences emerged for free resources, which were more important to respondents without a mental health history compared with those with a mental health history (P<.02), and consistent care from the same people over time on the digital platform (P<.01) and access to information on a lot of topics (P<.02) were more important for Hispanic and people of racial and ethnic minorities respondents than for non-Hispanic White respondents.

For both pregnant and postpartum survey respondents, the 2 least important features were care or content that fit with their culture and identity and proactive outreach from the digital resources. These features were selected as extremely important by less than 20% of respondents. In interviews, participants described experiences on digital health platforms that were enhanced by resources and support that were respectful of and relevant to their culture or identity. When specifically asked about the importance of digital health resources that “take their whole identity into account,” participants consistently interpreted this to be about race, ethnicity, and gender or sexual identity, and participants from non–historically marginalized populations tended to report that inclusion was not relevant for them. This emerged in comments such as “Personally, I’m a woman, heterosexual, so I fit in fine, more than fine” (Pilot P1) and “[Health care is] not hard for me to navigate as a straight, White woman” (Pilot P3). However, as they shared anecdotes about their use of digital health platforms, they described experiences that were meaningfully inclusive (or noninclusive) related to other characteristics, such as their lifestyle (eg, family structure, work choices, and how they cared for their child) or religious practices. A participant appreciated that a nutritionist took the time to accommodate the fact that she kept kosher and “went above and beyond to research kosher supplements and make sure they were appropriate” for her to take (Pilot P2), and a working mother described her appreciation for resources that recognized her more limited availability to be with her children:
I’m probably in the majority, for most of the like, you know, bubbles you would fill in of race, gender, like I’m not a minority in any of those, but I think that everyone was just very respectful...maybe because they got to just like learn a little bit about me. [P3]

People of racial and ethnic minorities participants appreciated when there were racially diverse providers on the platform. Even though the participants did not necessarily choose to interact with a member of their ethnicity or race, they valued the presence of diversity:

I felt included throughout the process. I think even seeing the diversity of the people you can choose for a specific topic also helps just, like, feel more included, like just her background. I am Mexican American. So I was born in the US but Mexican background. And so just seeing the diversity on the team was important to me. [P8]

Discussion

Principal Findings in Relation to Prior Research

This mixed methods study found that the reasons why perinatal people use digital resources are consistent with previous research. During pregnancy, our survey and interview participants were highly interested in fetal development [10,11,16,18,20,21], reassurance about what is normal [5,16,17,38], and childbirth [10,11,20,21,39], although in our study, the latter topic was not among the top reasons for use. Unlike other studies, we found less interest in health and nutrition during pregnancy. During the postpartum period, our results were concordant with those of previous research with regard to users having a high interest in feeding, especially breastfeeding; developmental milestones; and the baby’s health [18,19,21]. Although interest in infant sleep is also common in the literature [18,19], we did not explicitly ask about this. However, this need is consistent with our overall finding that respondents were looking for information about caring for their baby and with anecdotes in our interviews that centered on infant sleep when participants discussed wanting to know what is normal, reaching out to digital providers at night, and receiving nonjudgmental support. We found an especially high interest in using digital resources for breastfeeding help among respondents with a mental health history (17/20, 85%), which is consistent with studies showing an association between maternal anxiety and breastfeeding outcomes [40-42]. Aligning with findings from content analyses of perinatal forums, top interests included baby-related topics and pain during pregnancy [17,19], but we found less interest in using digital resources for the health of the birthing person, and we did not ask about relationships (a common topic in web-based perinatal forums [17,19]) as this did not emerge as a key interest during the formative interviews. In addition to finding many of the same known reasons for using digital resources, we found that using these resources to help manage anxiety or depression was commonly selected as extremely important (55/147, 37.4% during pregnancy and 45/110, 40.9% during the postpartum period), especially for people of racial and ethnic minorities during pregnancy (21/35, 60%).

Our survey identified that, during the postpartum period, reasons for using digital resources related to the health of the birthing person were ranked lower than reasons related to obtaining information or help about the baby. This aligns with findings that mothers tend to focus on their babies and deprioritize their own needs [43]. However, there are physical and emotional challenges during the fourth trimester that often go unaddressed and can negatively affect maternal and child outcomes [44,45]. Contributing to underuse is a misalignment between the issues being addressed by providers and the birthing person’s needs, discomfort with discussing “embarrassing” issues with providers and feeling judged by them, and not knowing where to access information [43,46,47]. Digital health resources provide an opportunity for birthing people to be more in control of when, how, and for which needs they access self-care resources. Even though self-care may not be the primary priority for using digital health resources, making these supports available is valuable for those times when the birthing person is ready and interested in using them.

Emerging from both our surveys and interviews was the preference for using digital health resources to access providers. The characteristics that make digitally based providers helpful are differentiators from in-person care, with providers on platforms being reported as spending more time, focusing on issues and questions that matter to the user, and providing support quickly and conveniently [48-50]. Although study participants were explicitly asked about what is important to them when using digital health resources in general, they all had access to Maven, which offers access to a wide range of provider services. Study participants may have been more likely to express the importance of digitally based providers as they had the opportunity to use them during their perinatal stage. Users of digital health services who do not have access to digitally based providers may not identify them as important as they have not had the opportunity to experience their value. Although our results may not replicate those among digital health users who have not had access to these providers, this does not mean that digitally based providers are not an important part of what should be offered in digital health. Rather, it is an opportunity to meet a need that people may not realize they have.

Generally, the needs and preferences for perinatal digital resource use did not show much variation by perinatal phase (pregnancy vs postpartum), parity, demographics, or health characteristics. The least important reasons for use were the same during pregnancy and the postpartum period: help with healthy eating and being physically active (and, for the postpartum period, help with losing weight), connecting with a community of people in the same perinatal stage, and obtaining recommendations of things to buy to care for the birthing person or baby. Platform features were ranked in generally the same order during pregnancy and the postpartum period. However, there seems to be a difference in the overall importance of information during pregnancy versus the postpartum period, with reasons for using digital resources being selected as extremely important by a greater percentage of respondents during pregnancy than post partum. During pregnancy, the only reasons for using digital resources that differed by race were
non–pregnancy-related (help with managing mental health and physical health problems), which may reflect gaps in in-person care for people of racial and ethnic minorities [51,52]. In interviews, we found a difference in participants who said that care or content that fits with their cultural, racial, or ethnic identity was important to them, with people of racial and ethnic minorities describing the value of being able to access providers with a shared identity or who provided culturally relevant support, whereas non-Hispanic White participants tended to say that this was not relevant to them. We were surprised that we did not find this difference in the surveys. A possible reason for this discrepancy is that our survey analysis focused on “extremely important” responses, and it may be that culturally relevant support is important but not extremely important. This would align with comments in the interviews about liking that providers of the same race were available even if they did not choose to use these providers.

This study included a national sample from the United States. Although our sample underrepresented people who identify as Black and Hispanic and lower-income people, it includes a significant portion of sustained users of digital resources (based on their completion of questionnaires on the Maven platform at multiple time points throughout pregnancy and the postpartum period), which is valuable as they represent people who will often use digital resources and have the potential to be significantly helped and supported by them. However, future research should also explore these topics among lower-frequency users to identify and understand barriers or deficiencies in the resources that may be contributing to reduced use.

This study adds to existing research by linking survey responses to other pregnancy, health, and demographic data about the respondents. Psychological states and mental health, especially perinatal anxiety, in relation to digital use are understudied [20], as are race and ethnicity. This formative research found only minor differences in digital resource needs and preferences across several user characteristics, including perinatal phase, parity, demographics, and mental and chronic health conditions. If this finding holds up in additional research, it has implications for the design of digital resources as it suggests that it may be appropriate to cover the same high-level perinatal topics for users across these groups. However, additional research is needed to better understand whether the content within these topics needs to be tailored to different populations. For example, future research could find that, although it is appropriate to include resources on childbirth for all users, there may be a need for more information about birthing centers in resources targeting Black pregnant people. Additional insights that are needed to improve the relevance and impact of these resources include understanding the use of the internet by nonbirthing partners [20] and the impact of use on clinical outcomes and user satisfaction [51]. There is also an opportunity to develop a standardized set of methods and metrics for assessing perinatal digital resources [4].

Limitations
This study has some limitations. First, the generalizability of these results is limited as all survey and interview participants had access to private insurance, the sample had a relatively high income, and Hispanic and Black respondents were underrepresented. Second, there was a high amount of missingness within our race and ethnicity data from respondents selecting that they prefer not to provide this information (17/110, 15.5% and 35/147, 23.8% for postpartum and pregnancy survey respondents, respectively). This missingness may limit the reliability of our results regarding differences in digital needs by race. Finally, not all survey respondents completed the Maven questionnaires, and those who did not complete at least one Maven questionnaire were excluded from the analyses of the survey. This potentially biased the sample toward heavier users of digital resources, although this creates an opportunity to understand use needs among this segment of users.

Data Availability
Owing to the sensitive nature of the questions asked in this study, respondents were assured that raw data would remain confidential but may be shared upon request.

Conflicts of Interest
NH, AB, AA, and HRJ are employed by Maven Clinic. NH, AA, and HRJ have equity in Maven Clinic.

Multimedia Appendix 1
Digital health platform features identified as extremely important during pregnancy by parity, mental health status, race, and ethnicity (N=147).

[PDF File (Adobe PDF File), 50 KB - formative_v8i1e48960_app1.pdf]

Multimedia Appendix 2
Digital health platform features identified as extremely important during the postpartum period by parity, mental health status, race, and ethnicity (N=110).

[PDF File (Adobe PDF File), 49 KB - formative_v8i1e48960_app2.pdf]

References


Abbreviations

IRB: institutional review board
Acceptability and Feasibility of a Smartphone-Based Real-Time Assessment of Suicide Among Black Men: Mixed Methods Pilot Study

Leslie B Adams, MPH, PhD; Thomasina Watts, MSPH; Aubrey DeVinney, BA; Emily E Haroz, PhD; Johannes Thrul, PhD; Jasmin Brooks Stephens, MA, LPA; Mia N Campbell, MHS; Denis Antoine, MD; Benjamin Lê Cook, PhD; Sean Joe, PhD; Roland J Thorpe Jr, PhD

Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States
Centre for Alcohol Policy Research, La Trobe University, Melbourne, Australia
Department of Psychology, University of Houston, Houston, TX, United States
Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD, United States
Department of Psychiatry, Cambridge Health Alliance, Cambridge, MA, United States
George Warren Brown School of Social Work, Washington University at St. Louis, MO, United States
Department of Health, Behavior, and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

Corresponding Author:
Leslie B Adams, MPH, PhD
Department of Mental Health
Johns Hopkins Bloomberg School of Public Health
624 N. Broadway
Baltimore, MD, 21205
United States
Phone: 1 410 955 1906
Email: ladams36@jhu.edu

Abstract

Background: Suicide rates in the United States have increased recently among Black men. To address this public health crisis, smartphone-based ecological momentary assessment (EMA) platforms are a promising way to collect dynamic, real-time data that can help improve suicide prevention efforts. Despite the promise of this methodology, little is known about its suitability in detecting experiences related to suicidal thoughts and behavior (STB) among Black men.

Objective: This study aims to clarify the acceptability and feasibility of using smartphone-based EMA through a pilot study that assesses the user experience among Black men.

Methods: We recruited Black men aged 18 years and older using the MyChart patient portal messaging (the patient-facing side of the Epic electronic medical record system) or outpatient provider referrals. Eligible participants self-identified as Black men with a previous history of STB and ownership of an Android or iOS smartphone. Eligible participants completed a 7-day smartphone-based EMA study. They received a prompt 4 times per day to complete a brief survey detailing their STB, as well as proximal risk factors, such as depression, social isolation, and feeling like a burden to others. At the conclusion of each day, participants also received a daily diary survey detailing their sleep quality and their daily experiences of everyday discrimination. Participants completed a semistructured exit interview of 60-90 minutes at the study’s conclusion.

Results: In total, 10 participants completed 166 EMA surveys and 39 daily diary entries. A total of 4 of the 10 participants completed 75% (21/28) or more of the EMA surveys, while 9 (90%) out of 10 completed 25% (7/28) or more. The average completion rate of all surveys was 58% (20.3/35), with a minimum of 17% (6/35) and maximum of 100% (35/35). A total of 4 (40%) out of 10 participants completed daily diary entries for the full pilot study. No safety-related incidents were reported. On average, participants took 2.08 minutes to complete EMA prompts and 2.72 minutes for daily diary surveys. Our qualitative results generally affirm the acceptability and feasibility of the study procedures, but the participants noted difficulties with the technology and the redundancy of the survey questions. Emerging themes also addressed issues such as reduced EMA survey compliance and diminished mood related to deficit-framed questions related to suicide.
Conclusions: Findings from this study will be used to clarify the suitability of EMA for Black men. Overall, our EMA pilot study demonstrated mixed feasibility and acceptability when delivered through smartphone-based apps to Black men. Specific recommendations are provided for managing safety within these study designs and for refinements in future intervention and implementation science research.

International Registered Report Identifier (IRRID): RR2-10.2196/31241

(JMIR Form Res 2024;8:e48992) doi:10.2196/48992

KEYWORDS
Black men; suicide; ecological momentary assessment; feasibility; acceptability; mixed methods; smartphone; real-time assessment; suicide prevention; user experience; behavior; implementation; intervention; mobile phone

Introduction
Suicide among Black Americans is a critical public health priority that requires immediate attention. Studies have shown that suicide rates among this population are increasing and that Black men, in particular, are at a higher risk of suicide-related mortality [1,2]. It is crucial for public health efforts to prioritize suicide prevention in the Black community and address the systemic issues that contribute to mental health disparities [1,2]. Epidemiological data reveal that Black men die by suicide at rates 4 to 6 times greater than Black women and that suicide is the third leading cause of death for young Black men [2,3]. While nationwide suicide rates provide a broader perspective on this alarming public health issue, delving into state-specific data, such as in Maryland between 2016 and 2020, reveal a localized increase in suicide death for Black Americans, compared to their White counterparts and highlights the urgent need for tailored research priorities [4]. The COVID-19 pandemic further exacerbated these alarming trends, with Black Marylanders experiencing an uptick in suicide deaths during the initial lockdown periods in the United States [5]. These national and state-specific trends highlight the pressing need to address more robust approaches to suicide prevention, emphasizing the importance of ensuring equitable access to support and resources for all communities.

Mental health disorders, such as depression and substance use, are known risk factors for suicide among Black Americans [6,7], but recent research also highlights racism and associated contextual stressors as important, but underresearched, explanatory risk factors for suicidal thoughts and behaviors (STB) among Black adults [8-10]. Previous studies have demonstrated the significant impact of racism and associated daily stressors on mental health outcomes, including depression, anxiety, sleep disturbances, and posttraumatic stress disorder (PTSD), among Black individuals [11-15]. However, the emerging research highlights the need to examine the specific ways in which racism contributes to STB among this population. Further research in this area is crucial to identify effective prevention strategies that can reduce the impact of racism and mitigate the rising trend of suicide completion among Black adults.

Black Americans also face significant barriers in accessing mental health care due to stigma, systemic lack of access to services, and cultural mistrust [16-19]. Although Black men are approximately 30% more likely to report having a mental health condition compared to non-Hispanic White individuals [20], they also have significantly lower percentages of mental health-related visits prior to a suicide attempt than men of other racial and ethnic backgrounds and Black women [21,22]. In addition, traditional clinical records used to assess suicide risk and prevention often rely on distal and static measures of risk, such as prior diagnoses and family history, which may overlook the dynamic interplay of proximal factors that contribute to STB and patient-provider synergy. Consequentially, while Black Americans face considerable obstacles in accessing mental health care, these disparities are particularly pronounced among Black men, highlighting the urgency of addressing these disparities and adopting approaches to assess more proximal aspects of suicide risk and prevention.

Smartphone-based ecological momentary assessment (EMA) platforms are one such approach to collect dynamic, real-time data and encompass a range of different active and passive information, including, but not limited to, spatial trajectories (via GPS), physical mobility patterns (via accelerometer), social networks and social dynamics (via call and text logs and Bluetooth), and EMA surveys [23-25]. Understanding dynamic risk in suicide prevention efforts is important since decades of previous research have shown that single, static risk factors often add little to our understanding of who is at elevated risk and when [26]. More real-time data on STB may also provide key information about when to intervene with potential just-in-time interventions.

Although participant burden, noncompliance, and reactivity to the protocol measures have been cited as potential limitations to EMA approaches, studies using this methodology to assess STB demonstrate a favorable median response rate of 70%, suggesting feasibility [25,27-30]. These studies highlight the potential of EMA as a tool for capturing the complexity and variability of STB, ultimately aiding efforts to improve mental health outcomes. Specifically, one systematic review of EMA studies found suicidality fluctuates considerably over short periods of time and that those with higher levels of overall suicidality also have more fluctuations [27]. This review also noted risk factors such as negative affect, hopelessness, burdensomeness, and sleep characteristics that impacted suicidality [27]. In a sample of psychiatric inpatients, a separate study discovered that the use of real-time data collections significantly enhanced the accuracy of predictions for suicide attempts post-discharge [31]. Despite the promise of this methodology for the study and prevention of suicide [25,29,32], there has been no study, to date, that uses this approach to assess experiences among Black men at critical periods for early
intervention [26]. Black men face daily societal and cultural stressors that may contribute to their heightened risk of suicide, including but not limited to, systemic and interpersonal racism, economic disparities, and limited access to mental health resources [8,33,34]. To address this gap, additional research is needed to clarify the efficacy of EMA monitoring as a suicide prevention tool for this at-risk population.

Methods

Eligibility and Recruitment Procedures

We recruited adult Black or African American men (18 years of age or older) with a lifetime history of suicidal ideation or attempt residing in Maryland counties where mobile crisis support was available. To be eligible, participants had to (1) own a smartphone and (2) not be present with active psychosis or cognitive deficits. We used two recruitment approaches: (1) providing recruitment information to eligible, active patients via the Johns Hopkins Health MyChart, the web-based patient portal of the Epic electronic medical record system that allows for health communication between patients and health care providers and (2) clinician referral. Participants who met the eligibility criteria through either of these approaches were referred to the study coordinator and were required to complete a screening survey to confirm their interest and eligibility.

Ethical Considerations

The study was approved by the institutional review boards of Johns Hopkins Bloomberg School of Public Health (IRB 00013672). All participants were provided detailed information about the study procedures and expectations, risks, and benefits as part of the informed consent process conducted by a research coordinator.

Baseline Survey

We asked enrolled participants to complete a brief baseline assessment including demographics and psychosocial measures associated with affective, gender, and race-specific factors associated with STB, such as anger, sadness, attributional style, and racial identity. A complete list of baseline measures is described elsewhere. During the informed consent process, participants received an overview of the MetricWire smartphone app from the study coordinator. The study coordinator also determined the county in which the participant resided, in order to align our safety protocol with mobile health crisis units in their area. After completion of the baseline survey, we asked participants to download the MetricWire app onto their personal smartphones for the duration of the study. The MetricWire app is available for both iOS and Android smartphone platforms at no cost in the Apple App Store (Apple Inc) or Google Play Store (Google), respectively. Examples of the user interface of the MetricWire app are presented in Figure 1.

EMA and Daily Diary Data Collection Procedures

This study used the MetricWire app to deliver daily EMA surveys to participants at 4 time points between 10 AM and 6 PM. Each of the prompts was designed to occur at least 2 hours after the previous prompt and included 3 push notification reminders at 20, 40, and 55 minutes after the initial prompt. If the participant did not complete the EMA survey within 60 minutes, it was marked as incomplete. The EMA surveys were designed to be brief and take no more than 3 minutes to complete to reduce respondent burden. To capture instances of racism-related stressors occurring outside of the random EMA survey prompts, participants were allowed to record event-driven entries detailing their daily experiences (see Figure 2).

Figure 1. Screenshot of suicidal ideation and intensity questions on the ecological momentary assessment user interface. CSSR: Columbia Suicide Severity Rating Scale.
We administered a brief daily diary survey via the MetricWire app once per day to assess participants’ everyday experiences, including sleep-related impairment and quality, as well as their daily experiences with racism-related stress. The daily diary was prompted each day at 8 PM. This survey was not designed to capture momentary instances but rather to provide a more comprehensive picture of participants’ daily experiences.

**Exit Interview**

After the 7-day data collection period, each participant underwent a qualitative semistructured exit interview conducted by the study team. The interviews probed participants on issues such as question difficulty and clarity, potential revisions to question prompts, and overall satisfaction with the study protocol and EMA surveys. Verbatim transcripts of the recorded interviews were produced and analyzed to evaluate the feasibility and effectiveness of the study protocol.

**Participant Incentives**

To encourage higher EMA survey completion rates, participants were eligible to receive up to US $110 throughout the study duration. The incentive was incrementally phased, with participants receiving US $10 after completing the baseline survey, US $20 for completing 20% (7/35) of all surveys, US $50 for completing 50% (17.5/35) of all surveys, and US $80 for completing 80% (28/35) of all surveys. We offered an additional US $20 for the completion of the exit interview.

**Safety Protocols**

Upon enrollment, we provided participants with a document containing information about local and national mental health resources, including suicide crisis hotlines, to support their well-being throughout the study. To ensure participant safety during the data collection period, we implemented a 3-tiered safety protocol. Moderate risk, which is defined as any suicidal ideations (“Have you had thoughts of killing yourself?”), since the last assessment, but without any plan or intent, resulted in a notification to the participant guiding them to the online support groups and community-based mental health services and urging them to seek support. Elevated risk, defined as suicidal ideation with intent or a plan within the last 24 hours (“Have you planned out how you would do it?” or “When you thought about making yourself not alive anymore, did you think that this was something you might actually do?”), resulted in the same notification that was given to participants with moderate risk and participants will be asked if they would like to handle the matter themselves or if they would like us to contact a mobile crisis response unit in their area on their behalf. Acute risk, defined as suicidal ideation with an action since the last assessment (“Did you do anything to make yourself not alive anymore or kill yourself?”), resulted in a call to the participant’s closest mobile crisis response unit made by a member of the study team on the participant’s behalf. To protect participant privacy, data were not stored on their smartphones. Instead, survey response data were automatically synced to the MetricWire servers when participants were connected to the internet.

Figure 2. Screenshot of the respondent-driven racial stress recording option on the ecological momentary assessment user interface.
internet, and encrypted response data were stored until the next connection. These servers were continuously backed up.

To provide additional support, all research personnel interacting with participants received training in psychological first aid to assist with identifying any mental health needs. EMA responses were closely monitored by the study coordinator and regular updates were provided during biweekly research team meetings between January 2022 and May 2023.

Quantitative and Qualitative Data Analysis
To measure the acceptability and feasibility of the EMA method, we tracked four key metrics using Excel (Microsoft): (1) the proportion of eligible patients who enrolled in the study, (2) the percentage of completed EMA sessions out of the total number of scheduled sessions, (3) the number of safety-related incidents reported to mobile health crisis support teams, and (4) the average time taken to complete the EMA surveys.

Semistructured exit interviews were audio recorded and transcribed verbatim by the research team. The transcripts were then imported into Dedoose (version 8; Sociocultural Research Consultants, LLC) qualitative data analysis software for analysis. A deductive coding framework was used based on previous conceptual definitions related to implementation science. The research team independently reviewed the transcripts and generated a preliminary codebook based on a priori concepts. These concepts focused on prevailing definitions of acceptability, feasibility, adoption, and fidelity to the EMA methodology.

Results

Demographics
Participants ranged from 18 to 34 years of age, with an average age of 27 (SD 5.31) years. The majority (4/10, 40%) of our sample had completed a high school degree, followed by an associate or bachelor’s degree (n=2, 20%), while some completed high school (n=1, 10%), college (n=2, 20%), and a master’s degree (n=1, 10%). Participants lived across 3 counties in Maryland, including Baltimore, Anne Arundel, and Howard. Furthermore, participants identified as a spectrum of sexual orientations, including heterosexual (n=3, 30%), gay (n=2, 20%), bisexual (n=1, 10%), pansexual (n=2, 20%), and questioning (n=1, 10%). One participant chose not to disclose his sexual orientation.

Feasibility of EMA and Daily Diary
The MyChart recruitment service identified and sent study information to 744 active patients residing in Maryland. A total of 58 individuals completed our interest survey. Of those, 53 (91%) were attributed to the MyChart recruitment service. Three submissions were from clinician referrals and 2 were from community outreach events. Out of the 58 individuals completing the interest survey, 10 (17%) participants were enrolled in the EMA study. The primary reasons for exclusion after completing the interest survey were active psychosis (n=19) and no history of suicidal thoughts or behavior (n=10). A total of 7 individuals were excluded because they reported having no mental health care provider.

The 10 enrolled participants completed 166 EMA surveys and 39 daily diary entries. A total of 4 (40%) out of 10 participants completed 75% (21/28) or more of the EMA surveys. The average completion rate of all surveys was 58% (20.3/35), with a minimum of 17% (6/35), and maximum of 100% (35/35). A total of 4 (40%) out of 10 participants completed at least 5 out of 7 daily diary entries. The response rate for each day of the week ranged from 64% (3.2 surveys submitted on average; day 1) to 54% (2.7 surveys submitted on average; days 6 and 7). All 10 participants completed qualitative exit interviews at the conclusion of the study. There were no safety-related incidents that required mobile health crisis service response teams. On average, participants took 2 minutes and 5 seconds to complete EMA surveys and 2 minutes and 43 seconds to complete daily diary surveys.

Qualitative Findings From Exit Interviews
After coding, several key themes arose, including the study implementation, the iatrogenic effects of repeated suicidality assessments, and strategies to improve usability and effectiveness in suicide-related smartphone app development for Black men. We present key details and findings on these themes as follows.

Study Implementation
Table 1 shows representative quotes that illustrate participant perspectives on adoption, acceptability, and feasibility to the study protocol.
Table 1. Qualitative codes and selected quotations related to implementation of smartphone-based ecological momentary assessment study.

<table>
<thead>
<tr>
<th>Code</th>
<th>Key critiques or recommendations</th>
<th>Key quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceptability</td>
<td>“The only thing that I can really think of as far as that is the hours...Maybe they're asleep, maybe they're working etc., etc.” [Participant 1]</td>
<td>“I think that the first couple of questions are pretty intense. It makes sense. It's exactly the information that you'd be looking for.” [Participant 1]</td>
</tr>
<tr>
<td></td>
<td>“I thought it shouldn't have been the same questions. It got boring and repetitive and seemed like more of a hassle than something that was going to help me, because it's annoying now, because it's the same questions.” [Participant 5]</td>
<td>“I like how it did call my therapist, because I finally set up a meeting with her. But she, you know, was like, 'Hey, um, you know, come in contact with me, because I know you're not doing well.' And I was like, 'Yeah, that's fair.'” [Participant 2]</td>
</tr>
<tr>
<td></td>
<td>“I thought the questions were pretty straightforward.” [Participant 4]</td>
<td>“One of the things that was a little annoying was like oh, annoying now, because it's the same questions.” [Participant 5]</td>
</tr>
<tr>
<td></td>
<td>“I think for the most part, it was good, it was self-explanatory. It was easy to get to.” [Participant 5]</td>
<td>“It seemed like every time I tried to open the app, it would crash on me on. So, I was like, 'Oh, okay, that was convenient.'” [Participant 6]</td>
</tr>
<tr>
<td>Feasibility</td>
<td>“The wording was, yeah, as I was saying before, it was very, extremely clinical, which I mean, you're not really supposed to be beating around the bush as far as the stuff is concerned.” [Participant 1]</td>
<td>“Um, that was fine. I was, I'm not sure if you remember, but I was a bit wary beforehand. I was worried that it might be a little bit too many, but no, it was fine. The amount was perfectly okay. The timing was fine too. Yeah, the timing was definitely fine as well.” [Participant 1]</td>
</tr>
<tr>
<td></td>
<td>“It was good, because some days I will remember, I have to take the survey so much and such time but other days, I wouldn't to remember. I'm glad they, the prompts you know, were there.” [Participant 6]</td>
<td>“It's, it just felt strange that, well, yeah, unless they don't notice them or somehow it slips past their schedule. Maybe they're asleep, maybe they're working etc., etc.” [Participant 1]</td>
</tr>
<tr>
<td></td>
<td>“Sometimes you miss an alert and yeah and that would be like, all day, that was my main focus, like oh my God, I got to remember my survey or whatever.” [Participant 5]</td>
<td>“It was easy to get to.” [Participant 5]</td>
</tr>
<tr>
<td>Fidelity</td>
<td>“The only thing that I can really think of as far as that is the hours but as I had mentioned, before we even started this, anyone who's downloading this resource is probably doing it with the intention on actually utilizing it. So I can't really see someone actively like disregarding the surveys, unless they either don't notice them or, well, yeah, unless they don't notice them or somehow it slips past their schedule. Maybe they're asleep, maybe they're working etc., etc.” [Participant 1]</td>
<td>“I found myself being very, at least the first or second time, I felt myself being very, how am I supposed to answer this? It's, it just felt strange that, well, yeah.” [Participant 1]</td>
</tr>
<tr>
<td></td>
<td>“I think that the first couple of questions are pretty intense ones. So, what was interesting for me at first—so I was having like a very good week, so the questions were a little bit jarring almost, if you're like, in the middle of a good day, and then you get those questions.” [Participant 8]</td>
<td>“I normally don't go through racial stuff, like around the time that I was doing the, well, around the time that I was, you know, taking part in this study, it was out of work.” [Participant 1]</td>
</tr>
<tr>
<td>Fidelity</td>
<td>“One of the things that was a little annoying was like oh, it would crash on me a lot. So I would have to re-open it, but when I would re-open it, it would take me right back to the question that it crashed on me on. So, I was like, ‘Oh, okay, that was convenient.’” [Participant 4]</td>
<td>“Sometimes you miss an alert and yeah and that would be like, all day, that was my main focus, like oh my God, I got to remember my survey or whatever.” [Participant 5]</td>
</tr>
<tr>
<td></td>
<td>“it seemed like every time I tried to open the app, it wouldn't even open, like sometimes it...ugh like it was so frustrating, so I just wouldn’t try to do it for a while.” [Participant 6]</td>
<td>“It's, it just felt strange that, well, yeah, unless they don't notice them or somehow it slips past their schedule. Maybe they're asleep, maybe they're working etc., etc.” [Participant 1]</td>
</tr>
<tr>
<td></td>
<td>“It's, it just felt strange that, well, yeah, unless they don't notice them or somehow it slips past their schedule. Maybe they're asleep, maybe they're working etc., etc.” [Participant 1]</td>
<td>“It was easy to get to.” [Participant 5]</td>
</tr>
<tr>
<td></td>
<td>“Sometimes you miss an alert and yeah and that would be like, all day, that was my main focus, like oh my God, I got to remember my survey or whatever.” [Participant 5]</td>
<td>“It's, it just felt strange that, well, yeah, unless they don't notice them or somehow it slips past their schedule. Maybe they're asleep, maybe they're working etc., etc.” [Participant 1]</td>
</tr>
<tr>
<td></td>
<td>“I normally don't go through racial stuff, like around the time that I was doing the, well, around the time that I was, you know, taking part in this study, it was out of work.” [Participant 1]</td>
<td>“It was easy to get to.” [Participant 5]</td>
</tr>
<tr>
<td></td>
<td>“One of the things that was a little annoying was like oh, it would crash on me a lot. So I would have to re-open it, but when I would re-open it, it would take me right back to the question that it crashed on me on. So, I was like, ‘Oh, okay, that was convenient.’” [Participant 4]</td>
<td>“Sometimes you miss an alert and yeah and that would be like, all day, that was my main focus, like oh my God, I got to remember my survey or whatever.” [Participant 5]</td>
</tr>
<tr>
<td></td>
<td>“it seemed like every time I tried to open the app, it wouldn't even open, like sometimes it...ugh like it was so frustrating, so I just wouldn’t try to do it for a while.” [Participant 6]</td>
<td>“It was easy to get to.” [Participant 5]</td>
</tr>
<tr>
<td></td>
<td>“Lack of use due to app issues: “I didn't always have like issues as far as that's concerned, I didn't really use it very often, explicitly, because around the time that I was doing the, well, around the time that I was, you know, taking part in this study, it was out of work.” [Participant 1]</td>
<td>“It was easy to get to.” [Participant 5]</td>
</tr>
<tr>
<td>Appropriate  ness</td>
<td>Timing of prompts: “I'm not sure if you remember, but I was a bit wary beforehand. I was worried that it might be a little bit too many, but no, it was fine. The amount was perfectly okay. The timing was fine too. Yeah, the timing was definitely fine as well.” [Participant 1]</td>
<td>“I normally don't go through racial stuff, like around the time that I was doing the, well, around the time that I was, you know, taking part in this study, it was out of work.” [Participant 1]</td>
</tr>
<tr>
<td></td>
<td>“I normally don't go through racial stuff, like around the time that I was doing the, well, around the time that I was, you know, taking part in this study, it was out of work.” [Participant 1]</td>
<td>“I normally don't go through racial stuff, like around the time that I was doing the, well, around the time that I was, you know, taking part in this study, it was out of work.” [Participant 1]</td>
</tr>
<tr>
<td>Adoption</td>
<td>“I found myself being very, at least the first or second time, I felt myself being very, how am I supposed to answer this? It's, it just felt strange that, well, yeah.” [Participant 1]</td>
<td>“I normally don't go through racial stuff, like around the time that I was doing the, well, around the time that I was, you know, taking part in this study, it was out of work.” [Participant 1]</td>
</tr>
<tr>
<td></td>
<td>“I think that the first couple of questions are pretty intense ones. So, what was interesting for me at first—so I was having like a very good week, so the questions were a little bit jarring almost, if you're like, in the middle of a good day, and then you get those questions.” [Participant 8]</td>
<td>“I normally don't go through racial stuff, like around the time that I was doing the, well, around the time that I was, you know, taking part in this study, it was out of work.” [Participant 1]</td>
</tr>
</tbody>
</table>

aN/A: not applicable.
Overall, participants found the frequency and timing of the EMA survey prompts to be appropriate. One participant commented that the questions provided “exactly the information needed,” while another appreciated the reminder to contact their therapist. Participants generally found the questions straightforward and easy to understand, and the survey itself was considered self-explanatory and easy to access. Participant I noted that they were initially concerned about the number of prompts, but ultimately found them to be manageable:

I’m not sure if you remember, but I was a bit wary beforehand. I was worried that it might be a little bit too many, but no, it was fine. The amount was perfectly okay. The timing was fine too.

However, some participants reported feeling that the racial stress record and daily diary questions regarding discrimination may not have been relevant to their daily experiences.

I normally don’t go through racial stuff, like I’m in a neighborhood that is predominately Black. I see one White person a day, maybe. [Participant 5]

Our study identified some challenges related to technology issues that impacted the fidelity of our study implementation and, ultimately, participants’ ability to complete the surveys in a timely manner. Participants identified app crashes as a significant issue, which led to frustration and difficulty in completing the surveys.

One of the things that was a little annoying was like oh, it would crash on me a lot. So I would have to re-open it, but when I would re-open it, it would take me right back to the question that it crashed on me on. So, I was like, “Oh, okay, that was convenient.” [Participant 4]

This issue not only caused inconvenience but also resulted in participants potentially losing progress on their survey responses. Consequently, this problem may have discouraged some participants from continuing with the study.

Influence of Mood on EMA Compliance

The narratives shared by the participants highlighted the possibility of diminished effect resulting from frequent evaluations of suicidality, which could potentially exacerbate negative mental health outcomes. Participants’ accounts of their study experience suggest that depressed mood prior to a notification to complete a survey may impact compliance with suicide-related assessments. Overall, as stated by 1 participant, “during moments where I already wasn’t feeling too great, it exacerbated things a bit.” Further, another participant in the study described how their survey compliance can be influenced by their mood:

It’s sort of like a reflex for like, whenever something buzzes on my phone, and I’m like, not in a good mood, immediately, like, swipe it [get rid of the notification]

Participant 2 noted that repeated assessments of suicidality could increase their awareness of passive suicidal thoughts, stating, “Not really trigger me, but be like, oh, yeah, I am, like, passively suicidal 99% of the time.” Another participant noted in his interview that he did not complete the last day of the trial due to an anniversary of a loved one’s death:

It was an emotional day and I didn’t want to continue because my responses would have changed drastically. I didn’t want nobody to look at me differently...I didn’t want to put myself in an uncomfortable position.

This response further cements the idea that response rates may be influenced by low mood. Several participants provided feedback on the survey questions, suggesting that shorter surveys may be more effective in assessing suicidality. They noted that if they responded with a positive mood score, assessed by the Patient Health Questionnaire-2 item measure (PHQ-2), it likely indicated they were doing well and not feeling suicidal. Multiple participants suggested that gauging mood before asking directly about suicide may improve the effectiveness of the survey.

Those questions are pretty good. But I would probably, like I’d probably try and, like keep the survey shorter...So maybe like gauge the mood before asking like, “Hey, are you going to kill yourself?” Because if I’m at like, 3, then the answer is like, it’s no, it’s, it’s gonna be no, you know, because I’m in like, a pretty good place. [Participant 2]

Given our study’s focus on suicide risk assessment, a significant portion of the survey questions pertained to assessing proximal or momentary risk for self-harm. Unfortunately, these questions were predominantly framed from a negative or deficit-oriented perspective, such as “Have you thought about killing yourself since the last prompt?” Several participants pointed out that this focus on negative experiences was contributing to their mood states and recommended that the survey be balanced with more positive questions.

It seems like everything was about if I have a terrible day...there should have been some questions that show record of your good days, too. [Participant 5]

Recommendations to Improve Usability and Effectiveness

As part of the qualitative exit interview, participants were asked to provide feedback and suggestions for the future use of the app. In relation to the app’s accessibility, several participants suggested incorporating features that would promote ease of use. Specifically, participants mentioned that a dark mode would be beneficial, as the bright white background of the app was uncomfortable for some. In addition, participants suggested incorporating a text-to-speech feature that would read the survey questions aloud to make it easier for individuals with visual impairments or reading difficulties.

To encourage better participation and compliance, some participants suggested that a more flexible schedule would better align with their daily routines and lifestyles. Some participants pointed out that they are less inclined to fill out surveys when they are occupied with other social activities. As a result, participants suggested a schedule that can be customized according to their personal needs and circumstances.
I guess, maybe if I could like, sort of choose my own schedule...like put it at times where I know that I'm going to be like, on my phone, or like, in my bed...I'm probably going to be in my bed not doing anything else, like probably playing a game on my phone. So I'll tell myself, “Hey, you have no excuse, fill this out so that they know you haven’t, like died.” But, if it's like, in the middle of like, going to the mall with my friends, I'm most definitely not going to stop and be like, “Hold on everybody, I have to fill this out” I'm just gonna forget about it. [Participant 2]

Participants provided various suggestions for enhancing the interactivity of the smartphone-based app, including incorporating a journaling component to allow them to document their thoughts in written form. Participant 2 described the potential benefit of such a feature, stating that it could provide a private space for them to express their emotions and thoughts more freely, without the fear of judgment or negative consequences.

I was thinking, maybe something where, you know, if you're not feeling that bad, you can just write down something...And knowing that it's going to medical professionals, I probably will be as kind of open as I am in my journal.

Additionally, other participants expressed interest in in-app features that would allow for personalized coping strategies, such as guided breathing exercises or links to mental health resources. These in-app features were seen as potential tools to promote engagement with the app and ultimately improve mental health outcomes. Finally, participants recommended that the survey include questions that allow for a more comprehensive and nuanced picture of their mood and well-being, and that these questions should be presented before questions pertaining to suicide risk to “ease into the heavier stuff,” as suggested by participant 1.

Discussion

Principal Findings

Our study aimed to assess the feasibility of using smartphone-based EMA to evaluate suicidality in real time among individuals in a high-risk sample of Black men. During the interviews, most participants generally reported that the frequency and timing of the prompts and reminders were suitable for the project and their daily lives. The results indicate that this approach shows promise for future research in this population. The results also suggested that while suicide-related smartphone apps have the potential to be effective tools for suicide prevention, they need to be developed and implemented with care. Even though the completion rate for EMA surveys was not optimal, we are encouraged by the high level of compliance with the daily diary entries. The study also did not pose any safety concerns, and the completion time for the surveys was found to be relatively short. Participants reported that the questions were generally relevant to their daily experiences and easy to understand. Additionally, the user feature that allowed for open-ended audio recordings was found to be useful in capturing daily experiences of racism, which is an important factor that could contribute to suicidality in this population. Finally, the qualitative findings from exit interviews suggest that the study procedures were perceived by participants to be acceptable and feasible.

Despite the promising results, the study revealed that the compliance rate was lower than in other studies not focused on Black men that leveraged EMA approaches to assess suicide risk [25,28]. One possible reason for this finding is that participants encountered technological issues and cited them as reasons for disengagement from the study. Participants noted redundancy in survey questions, which can lead to survey fatigue and reduced engagement. This redundancy might have contributed to the lower completion rates observed in the study. Moreover, the frequency of prompts may have contributed to the overall low compliance rate. The participants recommended fewer prompts and more flexibility in timing to enhance compliance.

Our findings also highlight the need for careful consideration of the potential harms of repeated assessments of suicidality, which focuses on a negative framing of mental well-being, and the need for the development of personalized interventions to mitigate any negative effects. Specifically, participant narratives from our qualitative exit interviews suggest that individuals experiencing negative moods may be less likely to engage with suicide-related assessments delivered via smartphone app notifications. Additionally, respondents felt that the EMA survey questionnaire could be more balanced by including both questions that assessed suicide risk and positive emotions as protective factors. One such measure, the Reasons for Living Inventory (RLI), was developed to identify factors that serve as deterrents against suicidal behavior. This inventory assesses positive emotions that support an individual’s decision to avoid death by suicide, should such thoughts emerge. Including measures, such as the RLI, that ground Black men to life is vital for addressing mental health challenges and fostering well-being for this demographic. By understanding how to balance the benefits of suicide-related assessments with their potentially harmful impact, scholars will be more equipped to identify protective factors and develop tailored interventions that are more likely to be used by Black men.

Our findings have important implications for the use of EMA among Black men. Smartphone-based methods, such as EMA and daily diaries, offer a unique opportunity to assess and evaluate suicide risk in real time. This approach represents a significant advancement in capturing momentary phenomena that could aid in suicide prevention, such as (1) characterizing dynamic mood changes over a short time span, (2) untangling stressful and racialized daily experiences, and (3) transmitting critical health information to health care providers to support care management. By using this approach, researchers can gain a more nuanced understanding of the experiences of high-risk populations and provide timely culturally relevant interventions to prevent suicide [23,24,32,33].

Moreover, researchers should be aware of certain populations that may be cautious to participate in research, such as Black men, particularly when it involves the explicit tracking and reporting of their daily experiences. This hesitancy can be
rational, given the history of systemic exploitation of Black communities in biomedical research [36]. Therefore, it is important for researchers to acknowledge and address these historical injustices to build trust and facilitate participation among these populations. Only by taking these steps can researchers ensure that their work is ethical, respectful, and ultimately beneficial to the communities they seek to serve [19,37,38].

Smartphone-based suicide prevention is also increasingly becoming a viable approach for combatting rising suicide rates, especially among young adults. The prevalence of health-related internet searches via smartphones, as evidenced by the high percentage of people who are 18-29 and 30-49 years of age and who use their phones for medical information, suggests that smartphones can be effective tools for suicide prevention. Additionally, the fact that Black and Hispanic Americans use their phones for medical information more frequently than White non-Hispanic Americans suggests that smartphone-based interventions may be especially important in addressing suicide risk in communities that experience higher levels of stigma and barriers to accessing traditional health care resources [39]. By leveraging the widespread use of smartphones, suicide prevention efforts can reach a larger and more diverse population, potentially reducing the burden of suicide on individuals, families, and communities.

Limitations
There are several potential limitations to our study that should be noted. First, our primary recruitment method was through MyChart messaging, which may have limited the pool of potential participants to those who are more likely to have additional physical comorbidities. The recruitment approach may have influenced the response rate, since participants may not have had the same motivation to respond to an interest survey, compared to a direct referral from a medical professional. Second, all participants were required to own a smartphone, which may have excluded individuals who did not have access to this technology. Additionally, we excluded Black men with active psychosis, which may have further limited the diversity of our sample. Given the potential for overdiagnosis of psychosis in underserved communities [40], it is possible that we also excluded potentially eligible individuals. Our study was also geographically restricted to Black men receiving health care in Baltimore, Maryland, and its surrounding counties. Future studies should consider additional venues and settings to recruit Black men who are not engaged in psychiatric care, including but not limited to social media, advocacy groups, and peer-led and community-based organizations. Finally, our study involved a relatively small number of participants, with 9 eligible participants completing the EMA surveys. This limited sample size may not fully represent the diversity and complexity of experiences among Black men and can affect the generalizability of findings to a larger population. Future studies should expand from our preliminary findings with a sufficient sample size to statistically investigate predictors of low compliance and user uptake.

Conclusions
This study is the first to our knowledge to address critical gaps in suicide research by incorporating EMA to improve the care for Black men who are at risk of suicide. Using EMA may be an important tool to help stem challenges to the timely assessment of suicide among Black men, who comprise the largest percentage of deaths by suicide (81%) within the Black community [4]. Our findings highlight how acceptable and feasible this method is for this high-priority population, as well as potential approaches to improve its fit.

Our study provides critical insights into the use of smartphones to capture real-time data for assessing the mental and emotional health of individuals in high-risk clinical samples of Black men. Our results highlight the suitability of EMA using smartphone-based approaches for studying sensitive topics related to suicide in vulnerable populations. Furthermore, the findings shed light on the next steps for creating more equitable suicide prevention approaches, including identifying areas of missing data and the cultural acceptability of smartphone-based tools for health promotion. However, further research is necessary to expand these tools to assess structural racism and other racialized factors that influence Black men’s daily lives.

Acknowledgments
LBA was supported by a grant from the Bloomberg American Health Initiative at Johns Hopkins Bloomberg School of Public Health, from the National Institutes on Mental Health (K01MH127310) and the American Foundation on Suicide Prevention (YIG-0-001-19). RJTJ was supported by grants National Institutes of Health (K02AG059140) and (U54MD000214). EEH was supported through the National Institutes on Mental Health (K01MH226335). The MyChart Recruitment Service is a service supported by the Johns Hopkins Institute for Clinical and Translation Research. The Johns Hopkins Institute for Clinical and Translation Research is funded by the National Center for Advancing Translational Sciences through the Clinical & Translational Science Awards Program.

Data Availability
The data set generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.
References


Abbreviations

- EMA: Ecological momentary assessment
- PHQ-2: Patient Health Questionnaire-2
- PTSD: posttraumatic stress disorder
- RLI: Reasons for Living Inventory
- STB: suicidal thoughts and behavior
Acceptability and Feasibility of a Smartphone-Based Real-Time Assessment of Suicide Among Black Men: Mixed Methods Pilot Study

©Leslie B Adams, Thomasina Watts, Aubrey DeVinney, Emily E Haroz, Johannes Thrul, Jasmin Brooks Stephens, Mia N Campbell, Denis Antoine, Benjamin Lé Cook, Sean Joe, Roland J Thorpe Jr. Originally published in JMIR Formative Research (https://formative.jmir.org), 22.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Privacy Concerns About Sharing General and Specific Health Information on Twitter: Quantitative Study

Pouyan Esmaeilzadeh, PhD
Department of Information Systems and Business Analytics, College of Business, Florida International University, Miami, FL, United States

Corresponding Author:
Pouyan Esmaeilzadeh, PhD
Department of Information Systems and Business Analytics, College of Business
Florida International University
Modesto A. Maidique Campus 11200 S.W. 8th St
RB 261 B
Miami, FL, 33199
United States
Phone: 1 3053483302
Email: pesmaeil@fiu.edu

Abstract

Background: Twitter is a common platform for people to share opinions, discuss health-related topics, and engage in conversations with a wide audience. Twitter users frequently share health information related to chronic diseases, mental health, and general wellness topics. However, sharing health information on Twitter raises privacy concerns as it involves sharing personal and sensitive data on a web-based platform.

Objective: This study aims to adopt an interactive approach and develop a model consisting of privacy concerns related to web-based vendors and web-based peers. The research model integrates the 4 dimensions of concern for information privacy that express concerns related to the practices of companies and the 4 dimensions of peer privacy concern that reflect concerns related to web-based interactions with peers. This study examined how this interaction may affect individuals’ information-sharing behavior on Twitter.

Methods: Data were collected from 329 Twitter users in the United States using a web-based survey.

Results: Results suggest that privacy concerns related to company practices might not significantly influence the sharing of general health information, such as details about hospitals and medications. However, privacy concerns related to companies and third parties can negatively shape the disclosure of specific health information, such as personal medical issues ($\beta=-.43; P<.001$). Findings show that peer-related privacy concerns significantly predict sharing patterns associated with general ($\beta=-.38; P<.001$) and specific health information ($\beta=-.72; P<.001$). In addition, results suggest that people may disclose more general health information than specific health information owing to peer-related privacy concerns ($t_{165}=4.72; P<.001$). The model explains 41% of the variance in general health information disclosure and 67% in specific health information sharing on Twitter.

Conclusions: The results can contribute to privacy research and propose some practical implications. The findings provide insights for developers, policy makers, and health communication professionals about mitigating privacy concerns in web-based health information sharing. It particularly underlines the importance of addressing peer-related privacy concerns. The study underscores the need to build a secure and trustworthy web-based environment, emphasizing the significance of peer interactions and highlighting the need for improved regulations, clear data handling policies, and users’ control over their own data.

(JMIR Form Res 2024;8:e45573) doi: 10.2196/45573

KEYWORDS

concern for information privacy; CFIP; peer privacy concern; PrPC; health information disclosure; Twitter; empirical study
**Introduction**

**Background**

Reports and analyses highlight that approximately 60% of health-related tweets contain links to health-related websites [1]. Twitter users frequently share health information related to chronic diseases, mental health, and general wellness topics [2]. Twitter is a popular platform for health-related conversations because it allows users to share their thoughts, experiences, and information in real-time [3]. This platform can be particularly useful for sharing information about health events, such as disease outbreaks or public health campaigns. In addition, Twitter can be used to connect with others with similar health concerns or interests and to access information from healthcare professionals and organizations [4]. Twitter is a social media platform with a character limit, which makes sharing detailed information about health issues difficult. Moreover, the information shared on Twitter may not always be accurate or reliable, as it is not always fact-checked or verified. However, Twitter is a common platform for people to share opinions, discuss health-related topics, and converse with a wide audience. According to a survey conducted by the Pew Research Center, 21% of Twitter users have used the platform to share information about a health condition [5]. The survey also found that 20% of Twitter users have followed a health organization or medical professional on the platform, and 15% have searched for information about a health condition on Twitter. Users often express their views about health policies, medical breakthroughs, health care services, and public health issues [6]. Health professionals, researchers, advocacy groups, and patients actively participate in these discussions, contributing diverse perspectives and sharing evidence-based information. This open and rapid exchange of ideas allows health information dissemination and facilitates conversations that can influence public opinion and policy decisions [7].

Sharing health information on Twitter raises privacy concerns as it involves sharing personal and sensitive data on a public platform [8]. Information privacy refers to individuals’ control over collecting, using, and disclosing their personal information. Regarding sharing personal health information, web-based information privacy refers to protecting sensitive health data from unauthorized access, secondary use, or disclosure [9]. It involves ensuring that individuals can make informed decisions about how their health information is shared and used and that appropriate safeguards are in place to protect the confidentiality and security of this information. The potential privacy risks associated with sharing health information on Twitter can be grouped into 3 reasons. First, potential identification and disclosure of personal information—sharing health information on Twitter can inadvertently lead to disclosing personally identifiable information. A study found that anonymized data from health-related tweets could be reidentified to reveal the identity of users [10]. Researchers were able to reconstruct personal health stories and connect them to specific individuals, highlighting the potential privacy risks involved. Second, data mining and analytics—third parties can analyze and use health-related tweets for various purposes, including targeted advertising or creating consumer profiles. Researchers analyzed tweets related to mental health and found that the content could be used to predict users’ self-reported diagnoses, medication use, and other personal information [11]. This demonstrates the potential for extracting sensitive health-related data from Twitter. In addition, a study used deep neural networks to identify personal health experience tweets, highlighting the potential for using Twitter as a data source for health surveillance studies [12]. Third, public disclosure of sensitive health information—sharing health information on Twitter might inadvertently expose individuals to public scrutiny and judgment. A study examined tweets related to mental health and found that users often disclosed personal experiences, symptoms, and treatments [13]. Although this sharing can provide support, it can also expose individuals to potential stigma, discrimination, or unwanted attention.

Previous studies suggest that individuals may be comfortable with sharing general information that is not sensitive on social media [14]. However, people may not be likely to share personal information, especially health-related data, owing to privacy concerns [15]. According to previous studies, privacy concerns can arise from companies’ information collection and use policies in the age of medical big data [16] and web-based social interactions that may threaten information privacy [17]. Twitter is reported as an important data set for vendors, researchers, and medical companies to collect health-related information [18]. Many medical companies collect health information and patient experiences from Twitter for big data analysis to find patterns for public health management [19]. Although big data collection and data mining techniques could help generate intelligence for monitoring public health issues, they can cause privacy concerns. Reports highlight that many Twitter users have experienced invasion of privacy owing to companies’ collection, sharing, and analytics practices that use information from their tweets, including private health information [20].

Although there are various studies of vendor-related privacy concerns [21] and peer-related privacy concerns [22], little is known about whether these 2 aspects of privacy concerns may collectively influence information-sharing behaviors. As privacy violations can be related to peers (such as inappropriate comments and unauthorized retweeting) and companies (sharing personal information with third parties), many studies are required to examine whether information-sharing disclosure can be affected equally by vendor-related and peer-related privacy concerns. In this study, we aimed to determine whether both aspects of privacy concerns (ie, concern for information privacy [CFIP] and peer privacy concern [PrPC]) can mutually influence users’ effects can dominate or overshadow the impact of the other. For instance, whether the nature of the relationships and contexts with 2 different information trustees (ie, vendors and peers) can influence information dissemination behavior. Thus, we argue that both aspects of privacy concerns should be considered in a model to better characterize information privacy on social media. Investigating the importance of privacy concerns related to companies and vendors (such as Twitter analytics) and web-based peers (such as retweeting a focal user’s health information without permission) in case of disclosing...
public and private health information on Twitter would be the main contribution that makes this study different from others. This argument is built based on 4 reasons. First, although the main interactions on social media are mainly peer oriented, vendors can still collect a lot of personal data (such as health information) without authorization and use it for unconsented purposes [23]. There have been instances of social media platforms, including Twitter, being used by organizations for health-related data mining and analysis [24]. Twitter data can provide valuable insights into health-related trends for health care organizations through analytics [25]. Researchers and companies (such as pharmaceutical manufacturers) have used Twitter data to track and analyze health-related trends, including disease outbreaks, medication use, and public health concerns [26]. For example, a study found that Twitter data could be used to track the spread of influenza and predict outbreaks [27]. Another study uncovered that Twitter data could be used to monitor adverse drug reactions and identify potential safety concerns [28]. Pharmaceutical companies have also used Twitter data to monitor medication use and patient experiences. For example, a study reported that Twitter data could be used to monitor patient experiences with antidepressant medications [29]. Moreover, it is common for organizations and vendors, including those in the health care industry, to monitor social media platforms to gather insights about consumer opinions, preferences, and trends [30]. Twitter, as a popular social media platform, has been used for these purposes [31]. Health care organizations and vendors may collect health-related information, such as discussions about medical conditions, treatment experiences, and patient preferences, from public Twitter profiles [19]. These insights can be valuable for marketing and market research purposes.

Second, although the primary purpose of peer-to-peer (P2P) interactions on social media is to maintain social connections with peers and there are no explicit business-to-consumer interactions, companies can still use social media analytics to investigate published health information. Companies can leverage various analytics tools to gather and find meaning and patterns in data collected from social channels to support business decisions (such as predicting the risk factors to manage public health) [32]. Third, individuals share a variety of information (such as about lifestyle, health status, chronic issues, and medication) on social media, which can be more sensitive than conventional e-commerce information (such as transaction records). Disclosing a wide range of information across social media platforms could raise concerns about whether companies and peers misuse the shared data (eg, health information). Fourth, Web 2.0, a fundamental technology supporting social media, mainly focuses on bilateral relationships between peers. However, it does not remove traffic between companies and social media users. Thus, peers can comment on conversations about a user’s health condition and share others’ personal health information on their own channels. In contrast, companies can use big data analytical tools to collect, use, or share users’ personal health data with third parties.

Study Objectives
The main objective of this study was to investigate the concept of information privacy concerns in the context of social media based on both vendor-related and peer-related aspects. To do so, we used the survey research methodology and Twitter as the empirical context. We also relied on theories discussing 2 dimensions of information privacy (ie, CFIP and PrPC) as the theoretical foundation of our proposed model. In this study, CFIP, emphasizing both consumer perspectives and company responsibilities, represented privacy concerns related to web-based practices of companies and vendors (such as collection and sharing of self-shared information), and PrPC referred to privacy concerns about losing control over digital communications and web-based interactions with peers. Thus, we suggest that information privacy on social media can be multidimensional, focusing on privacy violations associated with companies’ (vendors’) practices and sharing behaviors of peers. The integration of CFIP and PrPC can comprehensively present the entirety of privacy concerns about web-based health information. This study contributes to both theory and practice. We shed more light on information privacy conceptualization in the context of social media. This study also provides an interactive outlook and practical recommendations for handling privacy issues by explaining how web-based vendors and peers may cause privacy violations when dealing with health information (general and specific) shared over the web.

Variable Conceptualization, Theoretical Foundation, and Research Hypotheses

General and Specific Health Information
Individuals can use web-based channels to share general health information, such as information about treatments, medications, side effects, hospitals, medical costs, and healthy behaviors [33]. For instance, people are likely to tweet about general obesity-related topics, such as the relationship between fast food and weight gain [34]. Another study identifies general tobacco-related tweets (such as information about smoking, cigarette risks, and quitting) as the primary conversational data sets for health-related topics on Twitter [35]. Moreover, people can use tools such as Twitter to share specific health-related information, including past medical history, allergies, personal medications, private health issues, and signs and symptoms. For example, a study indicates that people disseminate information about diagnoses, advice based on personal experience, use of specific medications, side effects, negative reactions, and treatments on Twitter [31]. Another study highlights that people use Twitter to share their COVID-19–related symptoms and personal health issues during the early stages of the pandemic [36].

Sharing public and private health information can be valuable for web-based peers and affect their health-related decisions. General information can enable web-based users to find some facts about hospitals, physicians, and diseases. Disseminating specific information can share important insights and advice based on personal health conditions, medical treatments, care planning, and medical experiences with chronic diseases. General health information can be publicly available regardless of personal experiences. However, specific health information
can be unpleasant to share because it may contain more private information. As health information dissemination has 2 sides, questions still remain as to what dimensions of information privacy may strongly affect sharing behaviors on Twitter.

**CFIP Constructs**

There is evidence suggesting that companies use tweets to collect health information. For example, reports show that public health researchers use Twitter data to study the world’s health. A recent study indicates that the amount of textual health-related data, which could be personal, collected by various organizations is growing (especially during the COVID-19 pandemic) [37]. Another study argues that health care researchers and research companies have used social media data sources such as Twitter to study public health [19]. Owing to the importance of the Twitter database, the Centers for Disease Control and Prevention (CDC) designed a document to guide employees and contractors on using Twitter to disseminate health information and engage with individuals and partners [38]. A study indicates that companies increasingly use Twitter to share public health information and collect real-time health data using crowdsourcing methods [39]. Information privacy, which refers to people’s ability to control their information, is essential in e-commerce and social media [40]. Several studies explain the privacy concerns specific to the mobility data collection context [41]. Thanks to emerging technology (such as Web 2.0), protecting personal information has become a growing concern for web-based users. CFIP is a general concern about how organizations can use and protect consumers’ information [21]. CFIP explains concerns about organizations’ information collection practices, use policies, and access to consumers’ personal information [42]. Previous studies indicate that examining consumers’ concerns about how companies (vendors) may use their personal information significantly affects their willingness to engage in web-based transactions actively [43].

In this study, following most previous studies, CFIP is posited as a multidimensional construct with 4 dimensions to measure individuals’ concerns about organizations’ information privacy practices [44]. Collection pertains to individuals’ concerns about what web-based information is collected and whether such information is stored properly. Unauthorized secondary use explains individuals’ concerns about whether the information collected for a consented purpose may be unethically and illegally used for other purposes without obtaining authorization. Improper access implies individuals’ concerns about whether unauthorized people (entities) can access, view, and share their information. Finally, concerns about errors reflect whether individuals’ information is appropriately protected to minimize accidental or intentional errors [44]. Therefore, the multidimensional scale of CFIP reflects the complexity of individuals’ privacy concerns [21]. According to Stewart and Segars [40], CFIP is developed as a second-order construct with 4 reflective first-order factors. In this study, we also considered CFIP as a high-order construct with reflective factors. The logic behind conceptualizing this construct as reflective was that the privacy concerns related to companies are reflective of the 4 dimensions (ie, collection, unauthorized access, errors, and secondary use) and the expected interactions among them. Therefore, these dimensions can reflect the same theme and may covary.

Although sharing information on Twitter is more oriented toward interactions with web-based peers, privacy concerns about the collection and misuse of digitized health information by vendors and companies still remain significant. Previous studies provide strong evidence suggesting that web-based users of Twitter are concerned about several aspects of their information privacy, from collection of a lot of data to misuse [45]. Our study focused on individuals’ perceptions about general CFIP owing to policies and practices of vendors and organizations that may collect, access, and use health information shared on Twitter rather than concerns about a particular vendor. According to the four dimensions of the CFIP construct, individuals who demonstrate high privacy concerns believe that (1) a lot of health information is collected by organizations from users’ Twitter accounts, (2) such health information is not appropriately protected against possible errors, (3) various organizations may use health-related information on Twitter for other purposes without authorization (such as data mining, surveillance, research, and business intelligence), and (4) there is lack of visibility into accurate security measures to control who can access and use health information from tweets.

Thus, the CFIP construct can be extended to privacy concerns about a wide range of vendors and companies accessing and using tweets containing health information. This concern is not the same as privacy issues owing to interactions with a specific vendor in the context of e-commerce (such as retail platforms). In these conventional interactions, privacy concerns may focus on personal, factual information shared in web-based transactions and services (such as demographic information). However, CFIP in the social media domain deals with concerns associated with the following uncertainty: which organizations collect personal posts, which unauthorized entities can view and share information, why and how the information is used (for instance, data mining), and how the information is protected from internal and external errors and misuse. Therefore, we argue that CFIP cannot be ignored in examining information privacy in social media because users may not have direct relationships with organizations on these digital platforms, but they are still concerned about how their posts can be collected and misused by various companies.

Sharing general health information could indicate a user’s rich medical information and wealth of medical knowledge. In contrast, sharing specific health information can show that the user may want to contribute or seek informational and emotional support by disseminating personal experiences and medical history. However, when privacy concerns about the collection and misuse of shared data by organizations are not addressed, users are not likely to disseminate general or specific health information on Twitter. Moreover, we can expect that because specific health information is more sensitive and private, web-based users may generally become more cautious about sharing it. Therefore, we hypothesized the following:

- Hypothesis 1A (H1A): CFIP negatively influences general health information dissemination on Twitter.
PrPC Constructs

Owing to the nature of Web 2.0, users can communicate, create content, and share it via communities, social networks, and virtual worlds [46]. Web-based users can share a wide range of information and experiences on social media. The information can be objective (based on factual data) or subjective (based on personal interpretation, feelings, tastes, or opinions) [47]. The range can start with demographic information (eg, age, gender, and race); continue with political views, humanitarian opinions, and health information; and end with comments on others’ posts [48]. People can use different formats, such as text, pictures, and videos, to disseminate information. Peers are important components of social networks; however, they can threaten information privacy through inappropriate sharing behaviors and unintended consequences of web-based interactions [49].

Web-based transactions with peers on social media affect users’ decisions about whether they want to reveal their personal information (such as feelings and likes) and create an image consistent with their personal identity [50]. In this study, peers could be web-based friends who may have long-lasting and affect-laden connections with a user and any web-based users who interact through social media channels. Previous studies highlight the importance of PrPCs in the context of web-based interpersonal relationships where other peers can access and view a user’s web-based information [51]. Peer-related privacy refers to possible risks of privacy invasion because of direct and indirect web-based interactions with peers [17]. Social bots and fake and spam accounts can also raise privacy violation risks by potentially exposing several peers to a focal user’s posts using machine learning algorithms [52]. Previous studies indicate the threat of using social bots on social networks, increasing the likelihood of privacy breaches where even more private user data are exposed [53]. Understanding who can access web-based information (such as a post related to signs and symptoms of depression) and with whom such information is shared can significantly raise privacy concerns. For example, a study shows that sharing information with only selected friends in social networking services perceived higher control than sharing information with all friends [54].

Thus, information-sharing behaviors on social media may erode the ability of users to control their virtual space and personal boundaries. Leaving an inappropriate comment for a user who posted about seeking ways to lose weight, can increase privacy concerns about lack of control to maintain the privacy of their Twitter space. A study posits that managing the privacy of virtual territory refers to defining the level of access to and interaction others can have within a user’s territory (eg, allowing peers to see or comment on the post) [55]. Peers can also play a bilateral role in web-based social interactions. They can intentionally or unintentionally share a user’s personal health information with others and expose the user to others’ personal information that they might not like to view. The user may think that if others’ personal health information has been shared with me, my posts can also be revealed to others. Thus, communication privacy can significantly affect how individuals and relational parties share private information on social media [56].

A recent study defines PrPC as the sense of inability to control personal boundaries in web-based interactions owing to web-based peers’ behaviors [22]. They describe this term using 4 reflective dimensions: peer-related information privacy, psychological privacy, virtual territory privacy, and communication privacy. Peer-related information privacy denotes concerns about who can see what type of information and when and how such information is disclosed to other web-based peers. For posts shared by a user, the main concern is unauthorized access and secondary use of data by other peers. On Twitter, this can happen through retweeting and commenting. Peers can also initiate posts or conversation threads to disclose a user’s personal information without authorization. A privacy concern is about the accuracy of personal information shared by peers. Thus, peers’ sharing can be a source of private information leaks.

Psychological privacy explains the control over input information coming from others to shape feelings, opinions, and beliefs. Information sharing is 2-way traffic in social media (ie, from a user to peers and from peers to a user) [57]. As people are exposed to posts shared by celebrities, business magnates, politicians, and other web-based users, their behaviors and opinions are increasingly affected by input information from peers. Peers on social media can influence users’ behavior by applying social influence through public comments on posts [58]. Privacy concerns become more intense when users’ opinions and psychological independence are intentionally manipulated by social bots [59]. In this situation, users are not able to make a decision independent of other web-based peers’ ideas. Moreover, receiving a lot of unwanted information from peers may influence value systems, attitudes, identities, and choices.

Virtual territory privacy represents concerns about an individual’s inability to achieve control over other peers’ interactions with their virtual properties (such as Twitter accounts) and shared conversations (postings). Previous studies suggest that the sense of ownership and emotional attachment to personal territory can be generalized to the social media domain [60]. Similar to other personal belongings, virtual properties are seen as private. Thus, any unwanted addition to or revision of personal information can be considered as an intrusion, which may increase privacy violation risks [45]. Finally, communication privacy reflects an individual’s lack of control over how and when other peers can make direct web-based conversations. For example, peers may use various communication tools to engage individuals in a group conversation about potentially embarrassing or stigmatic health-related topics. Then, users may feel pressured by being involved in such undesirable conversations with unfamiliar people.

Individuals may become more likely to share general or specific health information on Twitter when they think it is useful for...
other web-based peers (eg, they can make better medical decisions). However, peer-related concerns may prevent them from disseminating such information. Peers are participants in social media and can freely collect and share information that is sometimes considered as unwanted interference. For instance, if peers retweet a post containing personal information about postsurgery recovery plans without authorization or tag a user who posted general educational content about HIV, these web-based interactions may violate privacy needs and raise privacy concerns. In return, users may change the pattern of health information dissemination and become more cautious in sharing medical facts or personal experiences. Thus, we formulated the following hypotheses:

- **Hypothesis 2A (H2A):** PrPC negatively influences general health information dissemination on Twitter.
- **Hypothesis 2B (H2B):** PrPC negatively influences specific health information dissemination on Twitter.
- **Hypothesis 2C (H2C):** PrPC has a more negative effect on specific health information dissemination than on general health information sharing on Twitter.

### CFIP and PrPC Are Privacy Concerns for Twitter Users

Although tweets are publicly accessible by default, users likely expect some degree of privacy and control over their personal health information shared on the platform. Previous literature has found that even when posting content publicly on social media, individuals still have privacy interests and concerns about how their data might be used or accessed [61]. General health information shared publicly on Twitter, such as mentions of hospitals, physicians, and common diseases, is not considered protected or private. However, more specific personal health details, such as past medical history, allergies, medications, and current symptoms, could reveal private information about an individual’s health status. Although these details may be shared publicly by default on Twitter, users likely still have privacy concerns about this content being widely disseminated or used without their consent.

The concepts of CFIP and PrPC capture these types of privacy concerns. Although users are voluntarily sharing health information publicly on Twitter, they may still desire control over how this data are accessed and used. CFIP reflects concerns about using or sharing personal health data by third parties such as researchers or companies without the user’s knowledge or permission. Even if users willingly post health information publicly, they may still desire control over how that data are collected, analyzed, or shared by entities such as researchers, pharmacies, insurance companies, and so on. PrPC represents concerns about controlling boundaries around health disclosures and limiting exposure to certain audiences, such as employers or insurers, who could misuse the information. Users must balance sharing personal details with managing social risks if the information reaches unintended viewers such as employers, family members, or friends. Thus, although Twitter data are technically public, users are likely to have nuanced privacy interests surrounding their health disclosures. Therefore, concepts such as CFIP and PrPC are useful for quantifying expectations regarding control, anonymity, and audience boundaries that persist even when posting health care–related content openly over the web.

### Difference Between the Conceptualization of CFIP and PrPC

#### Overview

We used an interactive approach to provide a holistic view of information privacy in the context of sharing health information on Twitter. Using this approach, this study actively engaged with the 2 aspects of privacy concerns (CFIP and PrPC) in a dynamic way, considering the interplay between them, as opposed to treating them as isolated, independent factors. Therefore, we examined how these 2 aspects of privacy concerns interact with each other and how this interaction affects individuals’ behavior on Twitter. It should be mentioned that the dimensions used for CFIP and PrPC may differ because of the different nature of the relationships and contexts involved. Although the underlying concept of privacy concerns remains the same, the specific dimensions or factors that contribute to CFIP and PrPC may vary owing to the distinct characteristics of vendors and peers as information trustees.

#### Role and Control

Vendors typically have a professional or business relationship with individuals, where they are entrusted with handling personal information for specific purposes (eg, health care providers and web-based retailers). In this context, individuals may be concerned about vendors’ control over their information: how it is collected, used, and shared; and the potential for data breaches or unauthorized access.

#### Trust and Reputation

CFIP dimensions often include factors related to trust and reputation, such as trustworthiness, perceived reliability, and credibility of vendors. As individuals rely on vendors to handle their personal information responsibly, dimensions related to trust and reputation become important for CFIP measurement.

#### Legal and Ethical Considerations

CFIP dimensions may also include factors related to legal and ethical considerations, such as compliance with privacy laws, informed consent, and transparency in data practices. Individuals may be concerned about whether vendors meet the legal requirements and ethical standards in protecting their health information.

In contrast, peers, who are individuals within an individual’s social network or community, may have different dimensions of privacy concerns. Social interactions, trust, reciprocity, and the potential for social consequences typically characterize peer relationship dynamics. Some factors that could influence PrPC dimensions include the following.

#### Social Norms and Expectations

PrPC dimensions may reflect concerns about social norms and expectations related to privacy within the peer group. Individuals may worry about how their health information might be perceived, shared, or used by their peers and the potential impact on their social relationships or reputation.
Social Influence and Peer Pressure

PrPC dimensions may capture the influence of peer pressure or the fear of negative social consequences. Individuals may be concerned about potential judgment, stigma, or discrimination based on their health information within their peer group.

Personal Boundaries and Intimacy

PrPC dimensions may include factors related to personal boundaries and the level of intimacy within peer relationships. Individuals may be concerned about the extent to which personal health information should be shared with peers and the potential impact on their privacy, autonomy, and self-disclosure.

Although the underlying concept of privacy concerns is present in both CFIP and PrPC, the dimensions may differ owing to the distinct characteristics and dynamics of the relationships involved. Thus, considering these differences when developing measurement instruments is important to accurately capture individuals’ concerns regarding privacy in different trust relationships.

Research Model

The model focuses on health information and Twitter (as the research context). There are a few critical differences in the privacy concerns around health information compared with other types of information. First, health information is considered to be very sensitive and private. It can reveal details about medical conditions, treatments, prescriptions, family history, and so on. Other types of information, such as social media posts or shopping habits, are generally not as sensitive.

Second, health information has strict legal protections such as Health Insurance Portability and Accountability Act in the United States and General Data Protection Regulation in the European Union. These laws place restrictions on how health data can be collected, shared, and used. Other information does not have the same level of legal safeguards. Third, health information could potentially be used to discriminate against people in areas such as employment, insurance, and so on. This type of discrimination is legally prohibited, but the risk remains owing to the sensitive nature of the data. Other data, such as social media posts, have less potential for this type of discrimination. Finally, breach of health information is considered very serious, given the sensitivity of the data. Strong security protections are needed, and breaches can carry heavy penalties. Breaches of other types of data may not have the same level of severity.

Regarding privacy on social media, there are some key characteristics of the concerns around Twitter compared with other platforms. First, most Twitter content is public by default, whereas other platforms such as Facebook allow more privacy controls. This can raise concerns about a lack of control over dissemination. Second, tweets are often archived and searchable indefinitely; therefore, there are concerns about permanent availability even for “deleted” content. Other platforms may have more ephemeral sharing. Third, the open nature of Twitter makes it easy for tweets to spread rapidly and become viral compared with platforms such as Instagram, where sharing can be more controlled. This raises concerns about loss of context and lack of containment. Finally, the ability to create anonymous accounts on Twitter is greater than that on platforms such as Facebook that require real identities. This raises concerns about harmful speech, misinformation, and so on.

We proposed the following research framework for disclosing general and specific health information on Twitter by integrating 2 aspects of information privacy concerns (Figure 1). As several studies may have found empirical evidence for the hypotheses proposed in this study, we need to clarify what is new in our study. First, this study integrated both aspects of privacy concerns for the first time in a model. Previous studies examined either privacy concerns related to companies’ practices with web-based information (CFIP) [62] or concerns related to the web-based behaviors of peers (PrPC) [63]. However, as mentioned in the previous section, individuals may be concerned about disseminating their health information on Twitter because companies’ collection practices and web-based peers’ behaviors could violate their privacy. In this study, we wanted to examine whether both aspects of privacy concerns (ie, CFIP and PrPC) can collectively change health-related information-sharing decisions or whether one can dominate the other. For instance, whether the nature of the relationships and contexts with 2 different information trustees (ie, vendors and peers) can shape information dissemination behavior. Second, as Twitter is considered as a rich database for collecting individual health-related information to examine sentiments and manage public health [64] and reports highlight that individuals may be concerned about web-based interactions with peers [65], Twitter would be the best research context to meet the goals of this study. Third, this study distinguished between general and specific health information. Thus, we could offer more insights about privacy concern levels and disclosure behaviors related to the 2 types of health information on Twitter. These 3 reasons can make our study different from previous studies in the privacy literature.

In addition, we controlled for several variables such as age, gender, education, Twitter experience, privacy violation experience, and misrepresentation of identity on Twitter. According to previous studies in the privacy concern domain, some demographics, such as age [66], gender [67], and education level [68], can affect people’s intention to disclose information on social media. Moreover, the impacts of these variables have been examined in previous studies investigating individuals’ perceptions about sharing eHealth-related information [69,70]. The effects of these variables are often controlled in previous studies in the field of information privacy threats [71]. Thus, we assumed that individuals of different ages, genders, and educational levels engage in various disclosure behaviors because they have diverse backgrounds, individual characteristics, and personal differences. Therefore, we considered these demographics to be control variables in the proposed research model.

Moreover, the effects of misrepresentation of identity, experience with technology, and privacy violation experiences are controlled in previous studies examining relationships between privacy concerns and self-disclosure [22,42]. Thus, it is believed that individuals with different privacy violation experiences, previous identity misrepresentation, and experiences with Twitter are more likely to demonstrate various
disclosure behaviors. Therefore, we treated these experience-related variables as control variables in our model.

**Figure 1.** Research model. 1A: hypothesis 1A; 1B: hypothesis 1B; 2A: hypothesis 2A; 2B: hypothesis 2B.

---

**Methods**

**Research Approach and Survey Development**

We administered a web-based survey questionnaire to achieve the defined objectives and test the proposed model and research hypotheses. The survey consists of 4 sections. In the first part, the purpose of the study is described clearly, and a qualifying question is used to select respondents. The question for filtering respondents attempts to screen individuals with a Twitter account. Thus, individuals without a Twitter account are excluded from data collection and analysis. In the second section, respondents are asked to express their perceptions about privacy concerns associated with companies and third parties, peer-related privacy concerns, and health information dissemination behaviors. In the third section, demographic questions (ie, age, gender, education, income, and race) are asked. Finally, the last section focuses on personal privacy experiences (ie, Twitter experience, privacy violation experience, and misrepresentation of identity).

Questions to measure each construct were adapted from validated instruments available in the existing literature. Slight changes in the wording were made to fit the context of this study. We adapted items to measure CFIP (as a second-order construct with 4 dimensions) from the study by Stewart and Segars [40]. Following Zhang et al [22], we also conceptualized and measured PrPC as a second-order reflective construct with 4 dimensions. Previously defined scales to measure general and specific health information disclosure were adapted from the study by Hsu et al [72]. Respondents rated all the measuring items included in the survey using a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Multimedia Appendix 1 shows the questions used in the web-based survey.

**Data Collection and Data Analysis**

Data were collected in April 2022 by uploading the questionnaire to Amazon’s Mechanical Turk (MTurk). MTurk is a crowdsourcing platform that enables researchers to access data from potential target samples to conduct a study. MTurk has been recognized as an acceptable web-based means for collecting individual-level data. Literature about health care analytics shows a growing number of studies using MTurk for health-related research [73]. Previous studies highlight that MTurk can measure individual perceptions in various domains, such as social media [74]. As the target population of this study was US citizens who use Twitter for web-based interactions, we limited the respondents’ location to the United States. Moreover, 2 attention-check questions were used to remove participants who chose answers without correctly replying to reverse-coded filler items [75]. The filtering questions were as
follows: (1) It does not bother me that my peers may try to influence me through comments on my health-related postings on Twitter and (2) I am not concerned that I have little control over who can start a health-related conversation with me on Twitter. We received 364 questionnaires and excluded 35 (9.6%) that were either incomplete or failed the response quality questions, resulting in 329 (90.4%) valid and usable responses. The average response time to complete the questionnaire was 12 minutes. The descriptive statistics for demographics were performed using SPSS (version 26; IBM). The research model was tested using AMOS (version 26; IBM) within the structural equation model framework.

**Ethical Considerations**

The institutional review board of Florida International University reviewed and approved the study (approval 112755). According to the institutional review board approval, written informed consent to participate in the study was obtained from all participants. Moreover, the data collected in this study were anonymous. We considered US $1 as an incentive for each respondent to participate in the study.

**Results**

**Instrument Validation**

We used confirmatory factor analysis to assess convergent and discriminant validity. Table 1 shows the results of the convergent validity test. The standardized factor loadings for all constructs exceeded 0.7, which is the acceptable range for factor loadings [76]. The composite reliability values and Cronbach α values were above the recommended value of .7, demonstrating the adequate reliability of the constructs [77]. All the values of average variance extracted (AVE) exceeded 0.5, which is the cutoff value [78]. These measures indicated the acceptability of the measurement model’s convergent validity.

Table 2 shows the discriminant validity of the constructs. All diagonal values (square roots of the AVEs) were >0.7 and greater than off-diagonal values (correlations) between any pair of constructs [79]. Thus, the discriminant validity requirements were satisfied for the research model.

Moreover, we checked the convergent and discriminant validity of the second-order constructs. The composite reliability, Cronbach α, and AVE values for CFIP were 0.91, .88, and 0.64, respectively, and these measures for PrPC were 0.94, .89, and 0.72, respectively. The correlation between the second-order variables (e.g., CFIP and PrPC) was 0.58. Finally, the square roots of the AVEs for both constructs were >0.7 and higher than the correlations between the constructs. These results confirm an acceptable convergent and discriminant validity for both second-order constructs in the model.
Table 1. Results of the convergent validity test.

<table>
<thead>
<tr>
<th>Constructs, subdimensions, and items</th>
<th>Standardized factor loading (&gt;0.7)</th>
<th>Composite reliability (&gt;0.7)</th>
<th>Cronbach α (&gt;0.7)</th>
<th>AVE (&lt;0.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concern for information privacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COLL&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.89</td>
<td>0.75</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>COLL1</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COLL2</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COLL3</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COLL4</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USU&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.90</td>
<td>0.79</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td>USU1</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USU2</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USU3</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USU4</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IAC&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.86</td>
<td>0.79</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>IAC1</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IAC2</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IAC3</td>
<td>0.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERR&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.89</td>
<td>0.84</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>ERR1</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERR2</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERR3</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERR4</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer privacy concern</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPC&lt;sup&gt;f&lt;/sup&gt;</td>
<td>0.88</td>
<td>0.84</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>PPC1</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPC2</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPC3</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPC4</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPC&lt;sup&gt;g&lt;/sup&gt;</td>
<td>0.89</td>
<td>0.88</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>CPC1</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPC2</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPC3</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPC4</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VTPC&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.89</td>
<td>0.87</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>VTPC1</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VTPC2</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VTPC3</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VTPC4</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer-related information privacy concern–SSIPC&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.87</td>
<td>0.82</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td>SSIPC1</td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSIPC2</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSIPC3</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constructs, subdimensions, and items</td>
<td>Standardized factor loading (&gt;0.7)</td>
<td>Composite reliability (&gt;0.7)</td>
<td>Cronbach α (&gt;0.7)</td>
<td>AVEa (&gt;0.5)</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>-----------------------------------</td>
<td>-----------------------------</td>
<td>-------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>SSIPC4</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer-related information privacy concern--PSIPCj</td>
<td>0.89</td>
<td>.87</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>PSIPC5</td>
<td>0.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSIPC6</td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSIPC7</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSIPC8</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHIDk</td>
<td>N/Ai</td>
<td>0.87</td>
<td>.84</td>
<td>0.64</td>
</tr>
<tr>
<td>GHID1</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHID2</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHID3</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHID4</td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHIDm</td>
<td>N/A</td>
<td>0.90</td>
<td>.88</td>
<td>0.69</td>
</tr>
<tr>
<td>SHID1</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHID2</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHID3</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHID4</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a AVE: average variance extracted.
*b COLL: collection.
*c USU: unauthorized secondary use.
*d IAC: improper access.
*e ERR: error.
*f PPC: psychological privacy concern.
*g CPC: communication privacy concern.
*h VTPC: virtual territory privacy concern.
*i SSIPC: self-shared information privacy concern.
*j PSIPC: peer-shared information privacy concern.
*k GHID: general health information disclosure.
*l N/A: not applicable.
*m SHID: specific health information disclosure.
Table 2. Results of the discriminant validity test.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Score, mean (SD)</th>
<th>CFIP-COLL&lt;sup&gt;a&lt;/sup&gt;</th>
<th>CFIP-USU&lt;sup&gt;b&lt;/sup&gt;</th>
<th>CFIP-IAC&lt;sup&gt;c&lt;/sup&gt;</th>
<th>CFIP-ERR&lt;sup&gt;d&lt;/sup&gt;</th>
<th>PrPC-PPC&lt;sup&gt;e&lt;/sup&gt;</th>
<th>PrPC-CPC&lt;sup&gt;f&lt;/sup&gt;</th>
<th>PrPC-VTPC&lt;sup&gt;g&lt;/sup&gt;</th>
<th>PrPC-PRIPC&lt;sup&gt;h&lt;/sup&gt;</th>
<th>GHID&lt;sup&gt;i&lt;/sup&gt;</th>
<th>SHID&lt;sup&gt;j&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFIP-COLL</td>
<td>4.03 (0.66)</td>
<td>0.81&lt;sup&gt;k&lt;/sup&gt;</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFIP-USU</td>
<td>4.01 (0.67)</td>
<td>0.83</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFIP-IAC</td>
<td>3.94 (0.77)</td>
<td>0.66</td>
<td>0.81</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFIP-ERR</td>
<td>3.85 (0.82)</td>
<td>0.67</td>
<td>0.75</td>
<td>0.81</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PrPC-PPC</td>
<td>3.92 (0.78)</td>
<td>0.32</td>
<td>0.17</td>
<td>0.43</td>
<td>0.81</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PrPC-CPC</td>
<td>3.86 (0.88)</td>
<td>0.24</td>
<td>0.33</td>
<td>0.36</td>
<td>0.68</td>
<td>0.81</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PrPC-VTPC</td>
<td>3.83 (0.86)</td>
<td>0.30</td>
<td>0.28</td>
<td>0.32</td>
<td>0.31</td>
<td>0.74</td>
<td>0.72</td>
<td>0.82</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PrPC-PRIPC</td>
<td>3.87 (0.81)</td>
<td>0.19</td>
<td>0.36</td>
<td>0.31</td>
<td>0.33</td>
<td>0.69</td>
<td>0.70</td>
<td>0.74</td>
<td>0.80</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>GHID</td>
<td>3.84 (0.85)</td>
<td>0.34</td>
<td>0.32</td>
<td>0.42</td>
<td>0.36</td>
<td>0.35</td>
<td>0.42</td>
<td>0.30</td>
<td>0.49</td>
<td>0.80</td>
<td>—</td>
</tr>
<tr>
<td>SHID</td>
<td>3.81 (0.91)</td>
<td>0.30</td>
<td>0.30</td>
<td>0.38</td>
<td>0.31</td>
<td>0.37</td>
<td>0.38</td>
<td>0.41</td>
<td>0.41</td>
<td>0.53</td>
<td>0.83</td>
</tr>
</tbody>
</table>

<sup>a</sup>CFIP-COLL: concern for information privacy–collection.
<sup>b</sup>CFIP-USU: concern for information privacy–unauthorized secondary use.
<sup>c</sup>CFIP-IAC: concern for information privacy–improper access.
<sup>d</sup>CFIP-ERR: concern for information privacy–error.
<sup>e</sup>PrPC-PPC: peer privacy concern–psychological privacy concern.
<sup>f</sup>PrPC-CPC: peer privacy concern–communication privacy concern.
<sup>g</sup>PrPC-VTPC: peer privacy concern–virtual territory privacy concern.
<sup>h</sup>PrPC-PRIPC: peer privacy concern–peer-related information privacy concern.
<sup>i</sup>GHID: general health information disclosure.
<sup>j</sup>SHID: specific health information disclosure.
<sup>k</sup>Italicization represents the square roots of the average variance extracted.
<sup>l</sup>Not applicable.

Respondents’ Characteristics

Table 3 shows the participants’ characteristics. The descriptive statistics demonstrate that respondents were fairly distributed across gender, where 52.9% (174/329) were men and 47.1% (155/329) were women. The age range was positively skewed, indicating that most participants were young, with a range between 25 and 34 years (155/329, 47.1%) being high, followed by the range between 35 and 44 years (102/329, 31%). Approximately half (178/329, 54.1%) of the respondents had undergraduate or graduate education levels, which aligns with previous studies highlighting that people with high education levels tend to search more often for web-based health information [80]. The annual income was fairly distributed, with income between US $60,000 and US $79,999 showing a high range (135/329, 41%) among the provided categories. Most respondents were White (174/329, 52.9%), followed by Hispanic and African American individuals. Approximately half (174/329, 52.9%) of the respondents reported using Twitter for 4 to 6 years. Overall, 52% (171/329) of the respondents indicated that they had a privacy violation experience at least once (for instance, their account was hacked), and 38.9% (128/329) mentioned that they tried to use a fake account on Twitter (at least once).
Table 3. Descriptive statistics (N=329).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>174 (52.9)</td>
</tr>
<tr>
<td>Women</td>
<td>155 (47.1)</td>
</tr>
<tr>
<td><strong>Age group (y)</strong></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>13 (4)</td>
</tr>
<tr>
<td>25-34</td>
<td>155 (47.1)</td>
</tr>
<tr>
<td>35-44</td>
<td>102 (31)</td>
</tr>
<tr>
<td>45-54</td>
<td>33 (10)</td>
</tr>
<tr>
<td>55-64</td>
<td>20 (6.1)</td>
</tr>
<tr>
<td>≥65</td>
<td>7 (2.1)</td>
</tr>
<tr>
<td><strong>Education level</strong></td>
<td></td>
</tr>
<tr>
<td>Elementary</td>
<td>10 (3)</td>
</tr>
<tr>
<td>High school</td>
<td>66 (20.1)</td>
</tr>
<tr>
<td>College</td>
<td>76 (23.1)</td>
</tr>
<tr>
<td>Undergraduate</td>
<td>105 (31.9)</td>
</tr>
<tr>
<td>Graduate</td>
<td>72 (21.9)</td>
</tr>
<tr>
<td><strong>Annual income (US $)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;20,000</td>
<td>23 (7)</td>
</tr>
<tr>
<td>20,000-39,999</td>
<td>59 (17.9)</td>
</tr>
<tr>
<td>40,000-59,999</td>
<td>49 (14.9)</td>
</tr>
<tr>
<td>60,000-79,999</td>
<td>135 (41)</td>
</tr>
<tr>
<td>80,000-99,999</td>
<td>46 (14)</td>
</tr>
<tr>
<td>≥100,000</td>
<td>16 (4.9)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>36 (10.9)</td>
</tr>
<tr>
<td>Asian</td>
<td>23 (7)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>66 (20.1)</td>
</tr>
<tr>
<td>Native American</td>
<td>13 (4)</td>
</tr>
<tr>
<td>White</td>
<td>174 (52.9)</td>
</tr>
<tr>
<td>Mixed</td>
<td>16 (4.9)</td>
</tr>
<tr>
<td><strong>Twitter experience (y)</strong></td>
<td></td>
</tr>
<tr>
<td>1-3</td>
<td>63 (19.1)</td>
</tr>
<tr>
<td>4-6</td>
<td>174 (52.9)</td>
</tr>
<tr>
<td>7-9</td>
<td>66 (20.1)</td>
</tr>
<tr>
<td>10-12</td>
<td>20 (6.1)</td>
</tr>
<tr>
<td>&gt;12</td>
<td>7 (2.1)</td>
</tr>
<tr>
<td><strong>Privacy violation experience (eg, being hacked)</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>171 (52)</td>
</tr>
<tr>
<td>No</td>
<td>158 (48)</td>
</tr>
<tr>
<td><strong>Identity misrepresentation (eg, using fake accounts)</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>201 (61.1)</td>
</tr>
<tr>
<td>No</td>
<td>128 (38.9)</td>
</tr>
</tbody>
</table>
Analysis of the Dimensions

When implementing second-order variables in a measurement model, there are 2 common approaches: the repeated items approach and the 2-step approach [81]. This study used a repeated items approach to measure reflective second-order constructs. In the repeated items approach, the indicators used to measure the second-order construct are included in the measurement model twice: once as indicators of the second-order construct and once as indicators of the corresponding first-order constructs [82]. This approach allows for a direct assessment of both the second-order and underlying first-order constructs in a single measurement model. The repeated items approach provides a holistic view of the measurement model by simultaneously assessing the second-order construct and its underlying dimensions [83]. Using the repeated items approach provides a more integrated perspective about how CFIP and PrPC are influenced by their respective first-order constructs. It allows for a direct examination of the relationships between the second-order construct and its underlying factors.

Both CFIP and PrPC are conceptualized as second-order reflective constructs, consistent with existing literature. A reflectively measured construct shares a common theme across subdimensions; the dimensions are expected to be highly correlated [84]. Table 2 shows that, consistent with reflective measurement, the 4 dimensions of CFIP (ie, collection, unauthorized secondary use, improper access, and errors) are highly correlated with each other. As expected, the 4 dimensions of PrPC (ie, psychological, communication, virtual territory, and peer-related information privacy concerns) are also highly correlated. Results show that the 4 dimensions of CFIP, as first-order factors, load significantly on the second-order construct. The loadings were 0.91 for collection, 0.80 for unauthorized secondary use, 0.95 for improper access, and 0.88 for errors. Therefore, the interaction among 4 dimensions reflects CFIP, which shares a common theme of losing control over information privacy owing to companies’ sharing behaviors. Furthermore, the 4 dimensions of PrPC also load significantly on the second-order construct. The loadings were 0.90 for psychological privacy concerns, 0.92 for communication privacy concerns, 0.86 for virtual territory privacy concerns, and 0.95 for peer-related information privacy concerns. Thus, interactions among these 4 dimensions represent PrPC, which exhibits a shared theme of privacy concerns about losing personal control owing to web-based peer behaviors.

Structural Model and Path Analysis

Consistent with privacy literature, we controlled variables such as age, gender, education, years of experience, privacy violation experience, and misrepresentation of identity in the structural model [42]. Findings demonstrate that when the control variables are present, the coefficients and $R^2$ change significantly. Specifically, when age ($β=-.12; P=.01$, education ($β=.19; P=.003$), and privacy violation experience ($β=-.58; P=.008$) are present in the model, they significantly influence health information disclosure. Thus, the findings confirm that young people with high education levels who have not experienced privacy violations are more likely to disclose health information on Twitter. However, no effects of gender, years of experience, and misrepresentation of identity were found on health information–sharing behaviors. We used the structural equation model technique to analyze the factors affecting health information disclosure on Twitter. The results of model fit indexes exhibit a good fit with the goodness-of-fit indexes ($χ^2_{353}=2.2;$ goodness-of-fit index$=0.84$; adjusted goodness-of-fit index$=0.81$; comparative fit index$=0.90$; normed fit index$=0.91$; incremental fit index$=0.90$; standardized root mean square residual$=0.03$; and root mean square error of approximation $=0.04$) where all indexes meet their recommended cutoff values [85]. Table 4 depicts the summary of path analysis for 4 hypotheses (ie, H1A, H1B, H2A, and H2B).

Figure 2 shows the standardized path coefficients of the structural model. Support is not found for H1A, which proposes that CFIP significantly influences general health information disclosure on Twitter ($β=−0.07; P=0.16$). In contrast, the findings support H1B by confirming that CFIP significantly attenuates sharing behaviors when disclosing specific health information on Twitter ($β=−0.43; P<0.001$). H2A, which posits that PrPC would directly affect the disclosure of general health information on Twitter, is supported ($β=−0.38; P<0.001$). The analysis also exhibits that PrPC negatively shapes the sharing of specific health information on Twitter ($β=−0.72; P<0.001$), and this significant relationship supports H2B.

Regarding H1C and H2C, an alternative model was created for each hypothesis, and the 2 relationships in that hypothesis were constrained [86]. Next, a 2-tailed $t$ test was used to compare the difference between the alternative and the original model. H1C posits a significant difference between the impact of CFIP on general and specific information–sharing behaviors. As the $t$ value was significant ($t_{165}=3.45; P<0.001$), we confirm that CFIP imposes a more negative effect on specific health information dissemination than on sharing general health information on Twitter. In addition, H2C proposes that people may disclose more general health information than specific health information owing to peer-related privacy concerns. The $t$ value was significant ($t_{165}=4.72; P<0.001$); thus, the effect of PrPC was more prominent in specific health information sharing than in disclosing general health information on Twitter.

Finally, the model explains 41% of the variance in general health information disclosure and 67% in specific health information sharing on Twitter. The $R^2$ scores suggest that the 2 aspects of information privacy concerns (ie, concerns about the web-based practices of companies and peers’ behaviors) can provide reliable explanatory power to predict the variance in sharing general and specific health information.
Table 4. Path analysis.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Path</th>
<th>Standardized coefficient, $\beta$</th>
<th>SE</th>
<th>Critical ratio</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A</td>
<td>CFIP$^a$–general health information disclosure</td>
<td>$-0.07$</td>
<td>0.04</td>
<td>1.54</td>
<td>Not supported</td>
</tr>
<tr>
<td>1B</td>
<td>CFIP–specific health information disclosure</td>
<td>$-0.43^b$</td>
<td>0.03</td>
<td>4.21</td>
<td>Supported</td>
</tr>
<tr>
<td>2A</td>
<td>PrPC$^c$–general health information disclosure</td>
<td>$-0.38^b$</td>
<td>0.05</td>
<td>5.37</td>
<td>Supported</td>
</tr>
<tr>
<td>2B</td>
<td>PrPC–specific health information disclosure</td>
<td>$-0.72^b$</td>
<td>0.05</td>
<td>7.12</td>
<td>Supported</td>
</tr>
</tbody>
</table>

$^a$CFIP: concern for information privacy.

$^b$Significance level, $P<.001$.

$^c$PrPC: peer privacy concern.

Figure 2. Model paths. H1A: hypothesis 1A; H1B: hypothesis 1B; H2A: hypothesis 2A; H2B: hypothesis 2B; *$P<.001$.

Discussion

Principal Findings

Information sharing is one of the most important objectives of social media. People use Twitter for conversation, and information sharing can initiate a web-based exchange of ideas about an issue. As health information is more sensitive than other types of personal information, disclosing such data can raise privacy concerns. Most previous studies have mainly focused on privacy concerns related to companies and vendors as they may collect and use individuals’ personal information for other purposes or may not properly protect the collected information [87]. Few studies have also explained privacy concerns related to the web-based behaviors of peers [22]. However, previous literature did not consider both sides of information privacy concerns in a model in the context of social media. Moreover, disclosure behaviors on social media can be contingent upon the type of health information owing to sensitivity levels. Few studies have examined the sharing...
behaviors based on the unique characteristics of general and specific health information [86].

Although both antecedents (CFIP and PrPC) have been examined separately in previous studies, this study’s findings could propose scientific novelty. The study differs from previous research in this field because we integrated 2 aspects of privacy concerns (eg, related to companies and peers) to investigate the disclosure of general and specific health information on Twitter. In this study, we examined whether both aspects of information privacy concerns can jointly influence sharing decisions related to health-related information or whether the effect of one aspect can be overshadowed by the other; for instance, whether the nature of the relationships and contexts with 2 different information trustees (ie, vendors and peers) can influence people to share their health information on Twitter. The findings indicate that privacy concerns related to companies play a more significant role in predicting specific health information than in predicting general health information. Privacy concerns related to companies’ practices reflect the collection and misuse of health information by vendors, such as concerns about using health information for data mining and research purposes [88]. Our findings demonstrate that Twitter users are more concerned about vendors and companies using or sharing their personal health information than general health information. Thus, the dimensions of CFIP (ie, collection, unauthorized secondary use, improper access, and errors) become salient only when people want to disclose specific health information (such as information about their chronic diseases, signs, and symptoms or personal health status). Sharing public health information about hospitals, medical costs, and medications is not significantly affected by concerns about how companies may use such information. A plausible justification is that general health information cannot reflect any personal information associated with an individual, and even if it is used for data mining or big data analysis, it will not violate the user’s privacy needs. Consistent with previous studies [89], individuals may deliberately want to share general health information on social media to increase public awareness and knowledge about a medical situation, such as COVID-19 symptoms and vaccination. Regardless of information accuracy or misinformation, users may engage in sharing their general medical knowledge and public information about treatment options with almost no or minor privacy concerns related to companies and vendors’ collection and use practices.

Our results also show that peer-related privacy concerns can significantly shape both general and specific health information sharing on Twitter. Although Twitter is not the same as web-based health communities designed to share health information, many individuals use tweets to share personal and public health information [90]. Web-based interactions with peers may affect their sharing behaviors as they may feel unable to control who can see, comment on, or exchange the health information they share on Twitter. The dimensions of PrPC (ie, psychological, communication, virtual territory, and peer-related information privacy concerns) are important factors in predicting how users may disclose public and personal health information. However, peer-related privacy concerns are more intense for sharing personal than general health information. This finding indicates that when a user wants to share public information about a disease (for instance, cancer, COVID-19, or HIV), they are still concerned about how peers relate such general information to their profile. This concern becomes more salient when the user decides to reveal personal health information about that disease, such as what treatments or medications they are using daily or what medical procedures they will undergo in the future. Previous studies associate sharing personal health information related to physical health problems or mental disorders on web-based P2P networks with stigma [91]. The more sensitive the health information, the more stigma is attached to sharing such information with peers. Thus, being judged by peers (close friends and other web-based users) because of sharing personal health information may prevent them from disclosing that content on Twitter.

Although results show significant impacts of both aspects of privacy concerns on sharing specific health information on Twitter, peer-related privacy concerns are leading factors in shaping personal health information disclosure, more so than privacy concerns associated with companies and third parties. This result confirms the importance of web-based interactions with peers on social media and how to deal with their sharing behaviors, such as commenting or tagging [60]. This finding implies the critical effects of Twitter friends and the circle of people who can see and share tweets about private health information. Individuals may be first concerned about the Twitter circle and how peers would react to the shared personal information about health status and then become worried about how many companies may access such data and how they would use or share them. Thus, secondary dissemination of personal health information by web-based peers through liking, reposting, retweeting, or commenting on posts is more challenging to the maintenance of privacy controls than secondary use of data or unauthorized access to such private information by companies and vendors. Our finding that peer-related privacy concerns have a strong impact on health information sharing compared with privacy concerns associated with companies and third parties offers a counternarrative to prevalent assumptions in digital privacy research. This could be attributed to Twitter’s highly interactive and public nature, which might accentuate peer-related concerns. Previous studies, mainly those conducted in the context of web-based shopping or general social media use, might have overestimated the role of concerns associated with companies and third parties owing to the commercial and private nature of these web-based activities.

**Theoretical Implications**

This study may offer some theoretical contributions. First, our findings have implications for information privacy research in social media by integrating the existing privacy concern perspectives. This study can open up the discussions through which privacy needs related to companies, third parties, and peer-related aspects can be addressed. Then, this comprehensive mechanism may strongly affect users’ sharing behavior patterns. Second, this study distinguishes the differences between sharing public and private health information on Twitter. Although disclosing specific health information may help share personal experiences related to various medical situations that could be useful for peers, it is more challenging than disseminating general health information. The findings demonstrate how
company-related and peer-related privacy concerns could shape the 2 types of information-sharing behaviors. Third, this study investigates the effects of information privacy mechanisms on health information sharing in the context of Twitter. The findings can promote discussions about health information disclosure in other P2P networks such as other social media platforms, virtual worlds, or Metaverse. Fourth, exploring the determinants of information sharing regarding different types of privacy concerns can expand our current understanding of knowledge acquisition. As sharing both general and specific health information on social media can contribute to people’s medical knowledge, addressing the barriers to specific health information sharing and removing the privacy challenges can significantly help the procedures of medical knowledge acquisition from web-based interactions with peers.

The study contributes significantly to the theoretical understanding of privacy concerns in web-based health information sharing. The evidence that peer-related privacy concerns influence more strongly than those related to companies and third parties highlights a potential oversight in theoretical perspectives. Current theories largely view companies as the predominant source of digital privacy concerns, and this may need re-evaluation. The results extend existing theories by emphasizing the role of peer interactions in privacy concerns, particularly in public and highly interactive web-based environments such as Twitter. This recognition of the social dimension of privacy concerns could be integrated into existing theoretical models to provide a more comprehensive framework for web-based privacy behavior. Furthermore, although our study is specific to Twitter and health information, the insights gained may have broad applicability. The potential role of peer-related privacy concerns could be a valuable area of exploration in other social media contexts and in sharing other types of sensitive information. Thus, our findings open up new avenues for theoretical exploration and suggest a need for further studies to fully understand the complexities of privacy behavior in the digital age.

Unlike other research approaches, such as experiments, observational data, or qualitative interviews to assess privacy concerns and information sharing, our study used a quantitative survey approach. This allowed us to capture data from a large and more diverse sample, providing a more robust and generalizable understanding of privacy concerns in web-based health information sharing. The strength of our quantitative approach lies in its ability to establish clear patterns and relationships among various factors influencing privacy concerns. This enabled us to derive a more comprehensive and systematic understanding of the factors that significantly influence privacy concerns and health information sharing on Twitter. In terms of comparison, our findings offer a novel perspective about the role of peer-related privacy concerns in shaping web-based health information–sharing behaviors. Previous studies have predominantly focused on company-related and third party–related privacy concerns. However, our study highlighted the paramount influence of peer-related privacy concerns, thus suggesting a reorientation of focus in subsequent research efforts in this area. Our study also provides quantifiable evidence about the relative influence of peer-related privacy concerns and such privacy concerns associated with companies and third parties on Twitter users’ health information–sharing behaviors. Such quantifiable insights could serve as valuable benchmarks for future studies seeking to measure and compare similar variables in different contexts or on different platforms. Our survey methodology, coupled with a comparative analysis of the findings, underscores the contribution of our study to the field, offering both nuanced insights and broad trends that enrich our understanding of privacy concerns and health information sharing on social media platforms.

**Practical Contributions**

This study also provides several practical and technical implications for promoting privacy protection on Twitter. To promote the sharing of specific health information, it is essential to address privacy concerns related to both companies and peers. However, addressing peer-related privacy concerns is vital to encourage the disclosure of general health information. This is because concerns related to companies and third parties do not significantly predict general health information sharing. Thus, a robust privacy policy cannot be developed regardless of information type. As the 2 types of health information require different ways of satisfying privacy needs, mechanisms and regulations facilitating general and specific health information sharing cannot be the same. Depending on the type of health information, it is essential to customize the ability of Twitter users to control their self-concept and meet different privacy protection requirements. General procedures and privacy policies to regulate the dissemination and use of personal posts are not sufficient to address the information privacy concerns. Twitter should allay users’ privacy concerns about sharing specific health information using advanced technology and management mechanisms. For example, Twitter can enable individuals to restrict access to their shared personal health information. Punitive regulations can be established for inappropriate behaviors (such as retweeting without consent) that may discourage sharing specific health information. All controlling mechanisms and privacy protection functionalities should be easy to understand and use and should not be an additional burden on the users.

As the 2 aspects (company-related and peer-related aspects) of privacy concerns manifest in several dimensions, different features can be developed to address the need for effective protection mechanisms. Twitter can add a new feature to tweets, enabling users to identify the sensitivity of posts related to health information. The content will be recognized as a private post if the sensitivity score (eg, calculated based on a scale ranging from 1 to 10) is more than average. Then, that post is automatically restricted from exposure to everyone, and users can share their thoughts and experiences only with a small crowd. Users can also define terms and conditions for peers who want to retweet sensitive posts. For instance, a “request for share” button can appear for each sensitive post, and peers can only share the posts when they get approval from the focal users. Given our findings, Twitter could introduce a feature that allows users to select the audience for their health-related posts, thereby addressing peer-related privacy concerns. They could also introduce a “Health Information” mode that automatically
applies high privacy settings for tweets marked as health related. In addition, given the significant role of knowledge in shaping privacy concerns, Twitter should consider educational campaigns or prompts to inform users about these features and the importance of privacy when sharing health information.

In May 2022, Elon Musk called for further investigation into the accuracy of spam and fake account estimates, which Twitter announced to be <5%. Fake and spam accounts could lead to undesirable social interactions with peers, unwanted peer-shared information, and an unpleasant web-based social environment. Twitter needs to use new procedures to detect spam and fake accounts and better control the functionality of Twitter bots to provide a more appropriate web-based social platform. This new mechanism could automatically limit the visibility of private posts containing highly sensitive health information to everyone, even to people who users follow. On the basis of the current Twitter privacy policy, people can mention who can reply to a specific tweet. However, it is hard to confirm who can actually see the posts because of bots and recommendation agents.

Social bots use computer algorithms to artificially create content and interact with people on social media [92]. Twitter bots can be manipulative and purposely change people’s attitudes and opinions about a topic [93]. For instance, bots can share posts with peers who are not following a user but usually read posts with health information content. The existence of bots may be useful for sharing general information but can be very harmful because it can increase exposure to private posts with sensitive health information. A plausible recommendation is to add a new category for sensitive content (such as specific health information) besides the photo, graphics interchange format (GIF), and poll categories. Then, users can create a new circle of people who are allowed to see, reply to, or share these sensitive posts. Users can also customize the configuration and limit the possible unwanted interactions by selecting who can see the posts but cannot share them. This small crowd can be saved for future use and can be easily modified later. The advantage of this new category is that people are notified to customize their Twitter circle depending on different content (eg, highly sensitive, semisensitive, and nonsensitive). For instance, a user can select everyone to see and comment on posts containing information about cryptocurrency, high-technology companies, or humanitarian issues. In contrast, the user can select a group of followers to see their thoughts about general health information and choose only a few close friends to see and comment on posts with sensitive health information.

Stringent privacy policies are required to enable Twitter users to limit who (ie, peers) can view, comment on, and share web-based content. People should be able to easily edit with whom they can share health information to exercise control over their personal digital information. Spambots on Twitter should also be controlled, modified, or filtered because they can involve potentially deceptive, harmful, or annoying activities. A more transparent policy is required to detect and deactivate invasive Twitter bots that can automatically like or retweet users’ postings without consent.

Finally, the insights from our study are not only limited to Twitter but also have implications for other social media platforms where users might share health information. Such platforms should recognize the significant role of peer-related privacy concerns and consider introducing similar audience control features. Health professionals and health-related organizations using social media for patient engagement should also be aware of these concerns and take steps to ensure that their communication respects patient privacy. Policy makers should consider our findings when developing regulations for health information sharing on social media to ensure that they address the most significant privacy concerns.

Limitations and Future Studies

Our study also has some limitations that can be considered as opportunities for future studies. First, a web-based survey through MTurk was used to collect data, which may be biased toward people familiar with crowdsourcing platforms. Future studies can use other data collection and sampling strategies, such as collecting data directly from Twitter. Second, we collected data from 329 Twitter users, which may not be a good representative of Twitter users. Next, studies can increase the sample size to reduce sampling bias and improve the generalizability of the findings. Third, we did not consider the effects of cultural dimensions (such as individualism, uncertainty avoidance, etc) on sharing health information on Twitter. It can be interesting for future studies to explore the effect of culture on disclosing different types of health information on social media. Fourth, our study tests a model to analyze health information sharing from the perspective of privacy concerns. However, there may be other essential variables. More studies are required to examine other factors inhibiting and promoting sharing behaviors on social media, such as reputation, incentives, trust, stigma, and social support. Fifth, this study did not examine the accuracy of the health information shared on Twitter or the risks of misinformation because it is not within the scope of this study. Future studies could expand upon our results to investigate the role of misinformation risks in information-sharing behaviors. Finally, our study focused on Twitter as a study context. We encourage future studies to extend the proposed model to other social media platforms (eg, Facebook, TikTok, and Instagram) where web-based interactions with peers are essential.

Conclusions

This study provides insights into health information sharing on Twitter from a privacy perspective. The findings propose that including CFIP and PrPC constructs can help in better conceptualization of information privacy concerns in the context of social media. The integration of these 2 aspects of information privacy can expand the discussion about internet privacy by addressing the privacy needs associated with the practices of companies, such as collection, unauthorized secondary use, improper access, and errors. It also considers psychological privacy concerns, communication privacy concerns, peers’ sharing behaviors, and territory privacy concerns related to peers in such interpersonal interactions. This interactive approach can provide a more comprehensive analysis of information privacy (related to web-based vendors and web-based peers) and adds...
a more substantial explanation of privacy needs on social media channels (such as Twitter). Privacy concerns may not always prohibit disclosure behaviors on Twitter; it depends on the type of health information. The findings demonstrate that peer-related privacy concerns are more salient to predicting general and specific health information sharing on Twitter than privacy concerns related to companies and third parties. The results could propose practical contributions by shedding more light on the negative impacts of web-based peer behaviors on losing personal control over digital communications and information access. Privacy policies should focus on companies’ practices, such as sharing users’ information with third parties for big data analytics. We suggest mitigating privacy concerns and promoting health information sharing on Twitter by creating policies that tailor privacy needs to the type of health information shared (ie, general or specific).

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Web-based survey—measuring items.

References


81. van Riel AC, Henseler J, Kemerer CM, Esmaeilzadeh JMIR FORMATIVE RESEARCH 2020.11.037


93. van Riel AC, Henseler J, Kemerer CM, Esmaeilzadeh JMIR FORMATIVE RESEARCH 2020.11.037

Abbreviations

AVE: average variance extracted

https://formative.jmir.org/2024/1/e45573

JMIIR Form Res 2024 | vol. 8 | e45573 | p.578
(page number not for citation purposes)
CDC: Centers for Disease Control and Prevention
CFIP: concern for information privacy
GIF: graphics interchange format
H1A: hypothesis 1A
H1B: hypothesis 1B
H1C: hypothesis 1C
H2A: hypothesis 2A
H2B: hypothesis 2B
H2C: hypothesis 2C
MTurk: Mechanical Turk
P2P: peer-to-peer
PrPC: peer privacy concern
Assessment of Qatar’s Health Care Community Call Center Efficacy in Addressing COVID-19 Pandemic Health Care Challenges: Cross-Sectional Study

Muhammad Atif Waheed¹, MBBS, MRCS, MRCGP, DPD, MBA; Lolwa Al Mannai¹, MD, ABCM; Hanan Khudadad¹, MPH; Jamil Alenbawi¹, PhD; Mariama Aminata Mansaray¹, LLB, MA; Samya Al Abdulla¹, MBBCh, ABFM

Primary Healthcare Corporation, Qatar, Doha, Qatar

Corresponding Author:
Muhammad Atif Waheed, MBBS, MRCS, MRCGP, DPD, MBA
Primary Healthcare Corporation, Qatar
Al Mina Street
Doha, 40000
Qatar
Phone: 974 33015895
Email: dratifwaheed@gmail.com

Abstract

Background: The global COVID-19 pandemic caused by SARS-CoV-2 created many unprecedented challenges for health care organizations worldwide, placing a great deal of strain on the health care systems, especially access to health care services. To address these challenges, Qatar established a centralized digital platform as a community call center, initially offering digital consultations via its hotline (number: 16000) and later expanding to include a COVID-19 vaccination hotline (number: 7077) for mass immunization.

Objective: This study aims to comprehensively examine the community call center’s operations and their significant role during the COVID-19 pandemic.

Methods: Retrospective data were collected from the Health Information and Technology Department of the Primary Health Care Corporation, Qatar, from March 29, 2020, to January 27, 2022. Data analysis for the hotline (number: 16000) focused on telephone and video call volumes, call response rates, abandonment rates, and call classification. In addition, data from the COVID-19 vaccination hotline (number: 7077) were analyzed for call volumes, call response rates, abandonment rates, appointment booking rates, confirmations, rescheduling, and cancellations.

Results: The hotline (number: 16000) received a substantial total of 429,212 calls, with 284,849 (66.37%) calls effectively answered. The average number of calls received per day during the study period was 640.61 (SD 470.53), and the average number of calls answered per day was 425.14 (SD 206.64). Notably, of the total 128,468 consultations, video consultations were conducted for 3810 (2.96%). Among the diverse call categories, diabetes mellitus (6284/84,299, 7.45%), prescriptions and medications (4709/84,299, 5.59%), hypertension (3874/84,299, 4.6%), vitamin D-related issues (3770/84,299, 4.47%), upper respiratory tract infections (2690/84,299, 3.19%), and COVID-19-related inquiries (2590/84,299, 3.07%) were most frequently addressed. For the COVID-19 vaccination hotline (number: 7077), an impressive total of 1,512,354 calls were received, with a 58.27% (n=881,305) call response rate. The average number of calls per day during the study period was 3828.74 (SD 2931.94), and the average number of calls answered per day was 2231.15 (SD 1496.02). Appointment booking accounted for 26.37% (265,721/1,007,596), appointment confirmation accounted for 10.24% (103,136/1,007,596), rescheduling accounted for 7.95% (80,124/1,007,596), and cancellations accounted for 1.6% (16,128/1,007,596) of the calls.

Conclusions: The findings of this research highlight the crucial significance of the community call center hotline (number: 16000) and the COVID-19 vaccination hotline (number: 7077) in effectively addressing the multifaceted challenges posed by the global COVID-19 pandemic. In Qatar, the community call center emerged as an indispensable and accessible centralized resource, facilitating streamlined digital consultations and vaccination appointments. The impressive call response rate highlights its operational efficiency, adeptly managing a diverse range of health-related issues. This study emphasizes the critical role of community call centers in health care emergency response, signaling their potential as invaluable assets for future preparedness and effective mitigation strategies during similar public health crises.
BACKGROUND
SARS-CoV-2, an enveloped single-stranded RNA virus with an epicenter in the Hubei province of China, has caused the current COVID-19 pandemic [1]. It initially appeared as an outbreak of an unknown cause of pneumonia and marked the seventh coronavirus outbreak [2]. Following its emergence, it rapidly spread globally, resulting in significantly high morbidity and mortality rates. At the time of writing this paper, on August 18, 2022, there were 589,680,368 confirmed cases of COVID-19 globally, with 6,436,519 deaths and a fully vaccinated population of 4,857,273,828 [3]. In Qatar, 7,355,038 doses of COVID-19 vaccine were administered until the date of writing this paper.

The COVID-19 pandemic poses many challenges to health care systems worldwide and places considerable strain on them. One of the main challenges is to respond to the crisis and simultaneously maintain the provision of essential health care services, which is crucial for a high-quality and resilient health care system. During the COVID-19 pandemic, many essential health care services were disrupted, including cancer screening, tuberculosis screening, HIV testing, outpatient services, and maternal and child health services [4]. The lack of health care workers, diversion of health care staff to COVID-19 management, cancellation of planned treatments, and risk of viral transmission during on-site patient visits have disrupted health care services [5]. It is essential to take necessary steps to curtail the spread of COVID-19 using available resources and the best use of digital technologies [6]. Different countries have responded differently to the COVID-19 crisis. In the United Kingdom, the National Health Services established a COVID Response Service, accessible via 111 phone lines, with the recruitment of 5000 call handlers and 1500 retired clinicians [7]. The Gulf Cooperation Council (GCC) countries implemented various measures in response to the pandemic, including lockdowns of major cities, airline suspensions, school and university closures, restrictions on social gatherings and sporting events, free health care provision, and active screening for COVID-19 [8]. In Saudi Arabia, the Umrah pilgrimage was suspended, and travel restrictions were placed on GCC citizens who had visited COVID-19–affected countries [9]. To protect public health as a national health strategy in Qatar, the Ministry of Public Health has provided digital solutions by providing remote access channels to health care services at the Primary Health Care Corporation (PHCC) and Hamad Medical Corporation in collaboration with the TASMU Smart Program Qatar, Ministry of Transport and Communication, Hukoomi, and Qatar Post, along with notable digital solution providers [10]. A hotline (number: 16000) was set up to provide 24/7 service to patients’ inquiries regarding COVID-19. In addition, the PHCC established an inbound community call center on March 29, 2020, accessible via the hotline (number: 16000), to provide digital virtual primary care consultations as an alternative to face-to-face health center visits. Physicians and nurses working in health centers who were deemed to be at a high risk for COVID-19, such as those with chronic conditions (hypertension, diabetes, ischemic heart disease, chronic kidney disease, pregnancy, immunocompromised state, etc.), were given the option of working in the community call center. On January 5, 2021, the community call center added a COVID-19 vaccination hotline (number: 7077) for booking, cancelling, and rescheduling COVID-19 vaccination appointments [11].

The community call center hotline (number: 16000) provides telephone and video consultation services to all registered patients from 28 health centers in Qatar. Individuals can access the call center through hotline (number: 16000) and book appointments for telephone or video consultation with physicians, dentists, and ophthalmologists from 7 AM to 11 PM. Initially, the calls were triaged by the nurses. For patients who are not registered with the health centers, such as visitors and single workers, the triage nurses direct them to the appropriate service. In cases of emergencies requiring immediate medical attention, patients are guided to dial 999 for ambulance services as appropriate. For nonemergency situations, visitors are directed to local health centers, whereas single male workers are referred to worker’s health centers (HC-21), which are operated by the Qatar Red Crescent Society (QRCS). A nurse-led telephone triage service is available from 11 PM to 7 AM, which signs posts patients to appropriate services [12]. Although primarily established for the COVID-19 pandemic, it provided all types of consultations, whether urgent or nonurgent. There were 7 physicians and 9 nurse stations. There was 1 ophthalmology station, whereas for the COVID-19 vaccination hotline (number: 7077), there were 22 stations. The staffing level of the community call center varied according to its operational needs during the study period.

Digital consultations offer several advantages for patients, providers, and health care systems [13]. Patients benefit from avoiding waiting in queues, reducing travel burdens [14], convenience, cost efficiency [15], accessibility [16], and high levels of satisfaction in primary care settings [17]. Satisfaction levels are particularly high in digital consultations, which include communicating with physicians, addressing patients’ concerns and queries, developing treatment plans, improving illness comprehension, and offering usefulness and reliability [5]. During pandemics, digital consultations provide an excellent alternative to traditional face-to-face consultations for patients [18]. Providers also experience advantages, including flexible working hours, the ability to work from anywhere; a reduced risk of infection; increased job satisfaction [13,19,20]; and less psychological distress, burnout [13], and sickness, which can be a burden on the organization. From an organizational perspective, digital consultations provide centralized operations with weekly statistics and demand forecasting, eliminating unnecessary patient visits to health centers and resulting in
smooth operations and reduced clinic congestion [13]. Moreover, they enable health care services in remote areas and offer an opportunity for service expansion whenever possible. In addition, they contribute to lower CO$_2$ emissions and cost savings [20].

Call centers can serve as central hubs to respond to public health emergencies by providing rapid information transfer to health care providers [21]. During the COVID-19 pandemic, telehealth call centers played a significant role in supporting rural community health workers in Uganda, facilitating prompt identification and referral of COVID-19 cases for appropriate care [22]. In South Korea, the telehealth system provided up-to-date information to callers, helping them protect themselves and others from COVID-19 effectively [23]. Similarly, China established psychological support hotlines to offer mental health assistance during the pandemic [24], and in Paris, France, health care workers benefited from psychological support services [25].

**Objective**

Despite the importance of call centers during the pandemic, there is limited information available on their operations and outcomes, especially in relation to the COVID-19 pandemic and the vaccination hotline. Existing literature does not include reports from GCC countries detailing the role and impact of community call centers during the pandemic. This study aims to fill this gap by examining call center operations; documenting their contributions; and analyzing call volumes, patterns, response rates, categories of calls, priorities, and problems or diagnoses. By investigating the effectiveness of community call centers in Qatar during the COVID-19 pandemic, this research seeks to provide valuable insights into their performance during public health emergencies.

**Methods**

**Study Design**

A cross-sectional study design was used to assess the community call center hotline operations and performance during the study period.

**Disease and Study Population**

This research mainly evaluated community call center services to handle the COVID-19 pandemic health care challenges, specifically focusing on their utilization and effectiveness. The study population consisted of patients seeking routine and urgent consultations and advice regarding COVID-19 and its vaccination.

**Location**

Data were retrospectively collected from the Health Information and Technology Department of the PHCC, Qatar headquarters, where the community call center is located.

**Time Frame**

Data for the community call center hotline (number: 16000) were collected from March 29, 2020, to January 27, 2022. Similarly, data for the community call center vaccination hotline (number: 7077) were collected from December 29, 2021, to January 27, 2022. This time frame encompassed critical phases of the COVID-19 pandemic and the subsequent mass vaccination campaign.

**Data Collection and Analysis**

This study involved meticulous analysis of 2 separate data sets: one pertaining to the hotline service (number: 16000) and the other concerning the COVID-19 vaccination hotline (number: 7077). The hotline service (number: 16000) data set was analyzed for various essential metrics, including call volume, percentage of calls answered or abandoned, call patterns, categories of calls, and specific health concerns addressed. Similarly, the COVID-19 vaccination hotline (number: 7077) data set was thoroughly scrutinized, considering call volume; call response rates; call patterns; and appointment-related information such as bookings, rescheduling, confirmations, and cancellations.

**Data Cleaning**

To ensure utmost data quality and reliability, a robust data cleaning process was meticulously executed. The 3-step approach encompassed vigilant screening of the databases to detect and address any suspicious features, precise diagnosis of faulty data to ensure accurate results, and appropriate treatment of identified discrepancies. Moreover, a separate copy of the original data was meticulously created in a new file to maintain data integrity during the cleaning process.

**Data Reporting**

Conforming to the highest standards of scientific reporting, this study adhered to the RECORD statement, an extension of the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement checklist. This ensured accurate and transparent reporting of the secondary data analysis, further enhancing the credibility of the study findings [26].

**Ethical Considerations**

This study obtained ethical approval from the PHCC Research Subcommittee, adhering to research ethics and ensuring participant protection (PHCC/DCR/2021/11/068). As the study used existing deidentified data, informed consent was waived in accordance with the approved protocol. Stringent privacy and confidentiality measures were implemented to safeguard the data. The study strictly complied with data protection regulations to maintain the confidentiality of sensitive information. No compensation was provided to human participants, as the study involved secondary analysis of deidentified data and did not involve direct interaction with participants. The research team took utmost care to handle the data responsibly and ethically throughout the study.

**Results**

**Community Call Center Hotline (Number: 16000)**

Table 1 shows the quarterly performance of the community call center hotline (number: 16000). During the study period, 66.37% (284,849/429,212) of the calls were handled, with the highest being handled in the third quarter (33,395/34,376, 97.15%), followed by the fourth quarter (34,543/35,885, 96.26%) of 2021.
The average number of calls per day during the study period was 640.61 (SD 470.53), and the average number of calls answered per day was 425.14 (SD 206.64).

Figure 1 shows the hotline (number: 16000) call volume patterns during the study period. The peak of calls occurred on April 19, 2020; April 7, 2021; and January 5, 2022.

Figure 2 shows the volume and pattern of the video consultations conducted during the study period. Video consultations accounted for 2.96% (3810/128,468) of the total consultations conducted. The number of video consultations conducted was highest in the beginning of the COVID-19 pandemic reaching to its peak in the first week of July 2020 and then gradually declined.

Table 1. Quarterly (Q) performance of community call center hotline (number: 16000).

<table>
<thead>
<tr>
<th>Time period</th>
<th>Total calls, n (%)</th>
<th>Total calls answered, n (%)</th>
<th>Total calls abandoned, n (%)</th>
<th>Calls/d, mean (SD)</th>
<th>Calls answered/d, mean (SD)</th>
<th>Calls abandoned/d, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1&lt;sup&gt;a&lt;/sup&gt; (n=3146)</td>
<td>3146 (100)</td>
<td>1906 (60.58)</td>
<td>1240 (39.42)</td>
<td>1048.67 (125.43)</td>
<td>635.33 (75.74)</td>
<td>413.33 (200.97)</td>
</tr>
<tr>
<td>Q2 &lt;sup&gt;a&lt;/sup&gt; (n=57,628)</td>
<td>57,628 (100)</td>
<td>37,732 (65.48)</td>
<td>19,896 (34.52)</td>
<td>633.27 (365.44)</td>
<td>414.64 (159.40)</td>
<td>218.64 (223.79)</td>
</tr>
<tr>
<td>Q3 &lt;sup&gt;a&lt;/sup&gt; (n=64,172)</td>
<td>64,172 (100)</td>
<td>37,224 (58.01)</td>
<td>26,948 (41.99)</td>
<td>697.52 (318.96)</td>
<td>404.61 (127.12)</td>
<td>292.91 (218.14)</td>
</tr>
<tr>
<td>Q4 &lt;sup&gt;a&lt;/sup&gt; (n=43,642)</td>
<td>43,642 (100)</td>
<td>27,834 (63.78)</td>
<td>15,808 (36.22)</td>
<td>474.37 (214.21)</td>
<td>302.54 (98.53)</td>
<td>171.83 (137.15)</td>
</tr>
<tr>
<td>2021</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1 &lt;sup&gt;b&lt;/sup&gt; (n=68,364)</td>
<td>68,364 (100)</td>
<td>35,070 (51.3)</td>
<td>33,294 (48.7)</td>
<td>759.60 (386.49)</td>
<td>389.67 (138.61)</td>
<td>369.93 (283.37)</td>
</tr>
<tr>
<td>Q2 &lt;sup&gt;b&lt;/sup&gt; (n=70,255)</td>
<td>70,255 (100)</td>
<td>49,831 (70.93)</td>
<td>20,424 (29.07)</td>
<td>772.03 (421.53)</td>
<td>547.59 (164.05)</td>
<td>224.44 (304.45)</td>
</tr>
<tr>
<td>Q3 &lt;sup&gt;b&lt;/sup&gt; (n=34,376)</td>
<td>34,376 (100)</td>
<td>33,395 (97.15)</td>
<td>981 (2.85)</td>
<td>373.65 (133.23)</td>
<td>362.99 (128.06)</td>
<td>10.66 (10.31)</td>
</tr>
<tr>
<td>Q4 &lt;sup&gt;b&lt;/sup&gt; (n=35,885)</td>
<td>35,885 (100)</td>
<td>34,543 (96.26)</td>
<td>1342 (3.74)</td>
<td>390.05 (258.57)</td>
<td>375.47 (212.67)</td>
<td>14.59 (59.67)</td>
</tr>
<tr>
<td>Q1 2022&lt;sup&gt;b&lt;/sup&gt; (n=51,744)</td>
<td>51,744 (100)</td>
<td>27,314 (52.79)</td>
<td>24,430 (47.21)</td>
<td>1916.44 (944.09)</td>
<td>1011.63 (220.61)</td>
<td>904.81 (764.96)</td>
</tr>
<tr>
<td>Total (n=429,212)</td>
<td>429,212 (100)</td>
<td>284,849 (66.37)</td>
<td>144,363 (33.63)</td>
<td>640.61 (470.53)</td>
<td>425.14 (206.64)</td>
<td>215.46 (311.91)</td>
</tr>
</tbody>
</table>

<sup>a</sup>March 29 until March 31, 2020.

<sup>b</sup>January 1 until January 27, 2022.
Figure 1. Call volume patterns of hotline (number: 16000).

Figure 2. Volume and pattern of video consultations over the hotline (number: 16000).

Figure 3. Problem-based summary of consultations over hotline (number: 16000). GERD: gastroesophageal reflux disease; URTI: upper respiratory tract infection; UTI: urinary tract infection.
Community Call Center Vaccination Hotline (Number: 7077)

Table 2 shows the quarterly performance of the COVID-19 vaccination hotline (number: 7077). The COVID-19 hotline (number: 7077) handled approximately 58.27% (881,305/1,512,354) of the total calls, whereas 41.73% (631,049/1,512,354) were abandoned during the study period. The average number of calls per day during the study period was 3828.74 (SD 2931.94), whereas the average number of calls answered per day was 2231.15 (SD 1496.02). The highest percentage of calls answered was during the third quarter of 2021 (112,445/118,372, 94.99%), while the lowest number of calls answered was during the first quarter of 2021 (180,205/311,364, 42.12%).

Figure 4 shows the volumes of calls and the pattern received by the COVID-19 vaccination hotline (number: 7077) during the study period. The highest number of calls was received around mid-May and November and the last week of December 2021.

Table 3 displays the quarterly call categories for the COVID-19 vaccination hotline (number: 7077). The highest number of appointments booked occurred in the fourth quarter of 2021 (86,924/1,007,596, 8.63%), followed by that in the second quarter (112,445/1,007,596, 8.07%). Similarly, the highest number of confirmed appointments occurred during the second (51,122/1,007,596, 5.07%) and fourth quarter (16,735/1,007,596, 1.66%) of 2021. The highest number of appointment cancellations occurred in the second (7028/1,007,596, 0.7%) and third quarter (2072/1,007,596, 0.21%) of 2021. Notably, the highest call volume identified for the workers’ health centers (HC-21) operated by QRCS occurred during the second quarter of 2021 (62,208/1,007,596, 6.17%), followed by the fourth quarter of 2021 (43,735/1,007,596, 4.34%). Calls from patients who did not meet the age vaccination criteria, as announced by the Ministry of Public Health of Qatar, were the highest in the second quarter of 2021 (57,081/1,007,596, 5.67%), followed by the third quarter of 2021 (8337/1,007,596, 0.83%).

Table 4 illustrates the distribution of call categories for the COVID-19 vaccination hotline (number: 7077) during the study period. Among the incoming calls, approximately 26.37% (265,721/1,007,596) were associated with appointment bookings, 10.24% (103,136/1,007,596) were for confirming existing appointments, 7.95% (80,124/1,007,596) involved rescheduling appointments, and 1.6% (16,128/1,007,596) were calls to cancel previously scheduled appointments. In addition, 14.8% (149,117/1,007,596) of the calls were identified as related to the workers’ health centers (HC-21) operated by QRCS, while 7.11% (71,600/1,007,596) of the calls received did not meet the age criteria for vaccination.

Table 2. Quarterly (Q) performance of COVID-19 vaccination hotline (number: 7077).

<table>
<thead>
<tr>
<th>Time period</th>
<th>Total calls, n (%)</th>
<th>Calls answered, n (%)</th>
<th>Calls abandoned, n (%)</th>
<th>Calls/d, mean (SD)</th>
<th>Calls answered/d, mean (SD)</th>
<th>Calls abandoned/d, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q4 2020&lt;sup&gt;a&lt;/sup&gt; (n=852)</td>
<td>852 (100)</td>
<td>297 (34.86)</td>
<td>555 (65.14)</td>
<td>284 (36.01)</td>
<td>99 (21.52)</td>
<td>185 (57.42)</td>
</tr>
<tr>
<td>Q1 2021 (n=310,842)</td>
<td>310,842 (100)</td>
<td>131,159 (42.19)</td>
<td>179,683 (57.88)</td>
<td>3453.80 (1735.19)</td>
<td>1457.32 (849.34)</td>
<td>1996.48 (1214.87)</td>
</tr>
<tr>
<td>Q2 (n=555,336)</td>
<td>555,336 (100)</td>
<td>356,259 (64.15)</td>
<td>199,077 (35.85)</td>
<td>6102.59 (2315.49)</td>
<td>3914.93 (885.59)</td>
<td>2187.66 (1788.70)</td>
</tr>
<tr>
<td>Q3 (n=118,372)</td>
<td>118,372 (100)</td>
<td>112,445 (94.99)</td>
<td>5927 (5.01)</td>
<td>1286.65 (887.29)</td>
<td>1222.23 (836.57)</td>
<td>64.42 (52.90)</td>
</tr>
<tr>
<td>Q4 (n=361,461)</td>
<td>361,461 (100)</td>
<td>190,959 (52.83)</td>
<td>170,502 (47.17)</td>
<td>3928.92 (3627.34)</td>
<td>2075.64 (1546.77)</td>
<td>1853.28 (2217.60)</td>
</tr>
<tr>
<td>Q1 2022&lt;sup&gt;b&lt;/sup&gt; (n=165,491)</td>
<td>165,491 (100)</td>
<td>90,186 (54.5)</td>
<td>75,305 (45.5)</td>
<td>6129.30 (2040.44)</td>
<td>3340.22 (650.37)</td>
<td>2789.07 (1792.58)</td>
</tr>
<tr>
<td>Total (n=1,512,354)</td>
<td>1,512,354 (100)</td>
<td>881,305 (58.27)</td>
<td>631,049 (41.73)</td>
<td>3828.74 (2931.94)</td>
<td>2231.15 (1496.02)</td>
<td>1597.59 (1790.57)</td>
</tr>
</tbody>
</table>

<sup>a</sup>December 29 until December 31, 2020.
<sup>b</sup>January 1 until January 27, 2022.
**Figure 4.** Volume and pattern of calls received by COVID-19 vaccination hotline (number: 7077).

**Table 3.** Quarterly (Q) call categories of COVID-19 vaccination hotline (number: 7077; n=1,007,596).

<table>
<thead>
<tr>
<th>Time period</th>
<th>Rescheduled, n (%)</th>
<th>Canceled, n (%)</th>
<th>Booked, n (%)</th>
<th>Confirmed, n (%)</th>
<th>&gt;1 appointment, n (%)</th>
<th>Worker’s health centers (HC-21), n (%)</th>
<th>Unmet vaccination age criteria, n (%)</th>
<th>Other n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q4 2020&lt;sup&gt;a&lt;/sup&gt;</td>
<td>81 (0.01)</td>
<td>14 (0)</td>
<td>0 (0)</td>
<td>78 (0.01)</td>
<td>13 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>68 (0.01)</td>
</tr>
<tr>
<td>2021</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>17,164 (1.7)</td>
<td>3514 (0.35)</td>
<td>42,559 (4.22)</td>
<td>13,195 (1.31)</td>
<td>823 (0.08)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>62,762 (6.23)</td>
</tr>
<tr>
<td>Q2</td>
<td>24,306 (2.41)</td>
<td>7028 (0.70)</td>
<td>81,331 (8.07)</td>
<td>51,122 (5.07)</td>
<td>1349 (0.13)</td>
<td>62,208 (6.17)</td>
<td>57,081 (5.67)</td>
<td>121,467 (12.06)</td>
</tr>
<tr>
<td>Q3</td>
<td>11,856 (1.18)</td>
<td>2072 (0.21)</td>
<td>20,442 (2.03)</td>
<td>14,737 (1.46)</td>
<td>656 (0.07)</td>
<td>16,006 (1.59)</td>
<td>8337 (0.83)</td>
<td>50,256 (4.99)</td>
</tr>
<tr>
<td>Q4</td>
<td>17,815 (1.77)</td>
<td>1943 (0.19)</td>
<td>86,924 (8.63)</td>
<td>16,735 (1.66)</td>
<td>240 (0.02)</td>
<td>43,735 (4.34)</td>
<td>4132 (0.41)</td>
<td>60,223 (5.98)</td>
</tr>
<tr>
<td>Q1 2022&lt;sup&gt;b&lt;/sup&gt;</td>
<td>8902 (0.88)</td>
<td>1557 (0.15)</td>
<td>34,465 (3.42)</td>
<td>7269 (0.72)</td>
<td>221 (0.02)</td>
<td>27,168 (2.7)</td>
<td>2050 (0.2)</td>
<td>23,692 (2.35)</td>
</tr>
<tr>
<td>Total</td>
<td>80,124 (7.95)</td>
<td>16,128 (1.6)</td>
<td>265,721 (26.37)</td>
<td>103,136 (10.24)</td>
<td>3302 (0.33)</td>
<td>149,117 (14.80)</td>
<td>71,600 (7.11)</td>
<td>318,468 (31.61)</td>
</tr>
</tbody>
</table>

<sup>a</sup> December 29 until December 31, 2020.

<sup>b</sup> January 1 until January 27, 2022.

**Table 4.** Distribution of call categories for COVID-19 vaccination hotline (number: 7077).

<table>
<thead>
<tr>
<th>Call category</th>
<th>Values (n=1,007,596), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Booked</td>
<td>265,721 (26.37)</td>
</tr>
<tr>
<td>Confirmed</td>
<td>103,136 (10.24)</td>
</tr>
<tr>
<td>Rescheduled</td>
<td>80,124 (7.95)</td>
</tr>
<tr>
<td>Cancelled</td>
<td>16,128 (1.6)</td>
</tr>
<tr>
<td>Worker’s health centers (HC-21)</td>
<td>149,117 (14.8)</td>
</tr>
<tr>
<td>Unmet vaccination age criteria</td>
<td>71,600 (7.1)</td>
</tr>
<tr>
<td>Other</td>
<td>318,468 (31.6)</td>
</tr>
</tbody>
</table>
Discussion

Principal Findings

According to the findings of this study, the community call center hotline (number: 7077) has played a pivotal role during the ongoing COVID-19 pandemic. These hotlines effectively answered 66.37% (284,849/429,212) of the calls for urgent consultation requests and 58.27% (881,305/1,512,354) of the calls for COVID-19 vaccination appointment inquiries. Notably, the community call center serves as a centralized and versatile resource for citizens and residents of Qatar, offering consultations for a wide range of cases, whether they are emergencies, urgent, or nonurgent, and enabling centralized operations for the mass vaccination program against COVID-19. It has provided a hub during challenging times when health care organizations have witnessed a rapid shift from face-to-face to virtual consultations.

The results of this study showed that after 4 weeks of the first COVID-19 case in Qatar, which was announced on February 29, 2020, the community call center was fully operating. The trajectory of the call volumes received is in line with the peaks of COVID-19 cases, with the first one occurring on May 30, 2020, the second occurring in the middle of April 2021, and the third occurring in the middle of January 2022. As a result, during the COVID-19 pandemic including the lockdown phases, the people of Qatar were able to receive the best medical care. This has also prevented the spread of COVID-19 while providing alternative solutions for patients to access primary health care services [27]. With the community call center, patients have been empowered in terms of flexibility and convenience to use virtual consultations as an alternative to traditional health care services.

The community call center’s diagnostic categories revealed its capacity to handle a diverse array of medical concerns, encompassing inquiries related to COVID-19; vaccination inquiries; acute problems; chronic diseases; medication and repeat prescription requests; requests for laboratory investigations; and various medical conditions related to ear, nose and throat, ophthalmology, dermatology, psychiatry, gynecology, gastrointestinal, neurology, cardiovascular, orthopedics, hematology, pediatrics, and endocrine and metabolic conditions. Interestingly, patients with diabetes mellitus accounted for the highest number of consultations, and virtual consultations have proved effective for managing type 2 diabetes with outcomes comparable with face-to-face consultations, as demonstrated in primary care settings in Australia [28]. Similarly, the management of hypertension via virtual consultations during the COVID-19 pandemic has offered a holistic approach to achieve hypertension control in the United States [29].

It is intriguing to note that video consultations experienced a significant surge in demand during the early phase of the COVID-19 pandemic, especially in the first week of July 2020, reaching a peak of 30 video consultations per day. However, following this peak, the use of video consultation gradually declined over time. Although there is some evidence indicating that video consultations were more sought-after during periods of higher COVID-19 cases, such as in early July 2020 and mid-April 2021, it is essential to highlight that this correlation is not consistently observed throughout the data set. Notably, there were instances of relatively low video consultation counts during times of higher COVID-19 cases, such as in mid-January 2022. The lower use of video consultations can be attributed to various factors, including patients’ preferences influenced by cultural and language barriers, their comfort or inclination toward telephone or face-to-face consultations, privacy and security concerns, and technical limitations. Cultural reasons play a role in shaping patient preferences; some individuals may feel more at ease and familiar with consultation methods such as telephone calls. The content, duration, and quality of video consultations might resemble those of telephone consultations, which could affect patient choices [30], aligning with the findings reported in existing literature. Interestingly, in the United Kingdom, video consultations constitute <1% of consultations in general practice [31]. This preference for other consultation methods, including telephone and face-to-face consultations, might be influenced by physicians’ perceptions that video consultations offer minimal advantages compared with the alternatives.

The COVID-19 vaccination hotline (number: 7077) played a crucial role in Qatar’s mass immunization program, facilitating services for booking, cancellations, and rescheduling initially based on professional group and age criteria. The program was launched on December 21, 2020, introducing the BNT162b2 (Pfizer-BioNTech) messenger RNA vaccine initially, followed by the mRNA-1273 (Moderna) vaccine 3 months later. Initially, priority was given to vaccinating frontline health care workers, individuals with severe or multiple chronic diseases, and those aged ≥70 years. Subsequently, the program was extended gradually by 1 age group at a time, along with selected professional groups, using age as the primary eligibility criterion throughout the rollout [32]. The data from the vaccination hotline revealed noteworthy call volume spikes during mid-May, mid-November, and the last week of December 2021, coinciding with new announcements related to the vaccination program’s expansion. However, the percentage of vaccination hotline’s answers was observed to be lower than that of digital consultation phones. This disparity in the response rates can be attributed to multiple factors. Periods of high demand, such as vaccination eligibility expansions or public announcements about vaccination campaigns, led to a surge in hotline calls, potentially affecting response times due to the increased call volume. Specific eligibility criteria for vaccination appointments also played a role, as some callers may not have met these criteria, resulting in redirection or absence of appointments. Technical issues, waiting times, and call congestion further influenced the hotline’s response rate. Notably, the third quarter of 2021 demonstrated a higher response rate for the vaccination hotline. This could be attributed to a relatively lower volume of calls received during this period, possibly influenced by factors such as expats traveling during summer vacations and the vaccination hotline’s capacity to meet the demand effectively. Understanding these dynamics is crucial for optimizing hotline performance and enhancing vaccination service accessibility. Both vaccines in Qatar have been found...
to elicit strong protection against COVID-19, prevent hospital admission, and reduce mortality [33]. Due to mass vaccination supported by the community call center and other measures taken by the Ministry of Public Health, Qatar’s mortality remained very low at 0.0016%, whereas globally, it was 1% at the time of writing this paper.

One of the most significant outcomes of the COVID-19 pandemic is that many outpatient appointments can now be managed efficiently via telemedicine without affecting patient care [34]. The COVID-19 pandemic has shown that health care workers can swiftly adjust to the new technologies required to use telemedicine [13]. In a study involving 23 primary care providers and 1692 patients, both providers and patients reported a desire to continue telemedicine visits after the pandemic in primary care settings in the United States [17]. Even before the COVID-19 pandemic, the telehealth business flourished, with a market value of >US $50 billion in 2019 and an anticipated growth rate of more than 9-fold over the next decade [35]. Virtual consultant jobs have also been advertised [36].

The significance of the community call center and vaccination hotline in providing consultations, vaccinations, and health care services during a challenging period demonstrates the potential and value of telehealth solutions in health care systems worldwide. By sharing the successful experiences and best practices of the community call center hotlines, this study can contribute to the enhancement of telehealth services and call centers globally. The lessons learned from operating the hotline can be invaluable for improving crisis response strategies and optimizing health care delivery during public health emergencies. By leveraging digital technologies, health care systems can enhance access to services, improve patient satisfaction, and manage public health emergencies effectively. Integrating call centers and telehealth into routine health care services and emergency response strategies can provide long-term benefits beyond the pandemic.

Limitations
This study has several limitations. The data presented in this study involved 3 peaks of COVID-19 that affected the overall percentage of calls answered by the hotline (number: 16000). The hotline answered 100% of all calls on several days of the week during the study period. Perhaps, the percentage of calls answered improved to 71.71% at the time of writing this paper. Similarly, for the COVID-19 vaccination hotline (number: 7077), a surge in the volume of calls due to a new announcement of rolling out of the vaccination program affected the overall percentage of the calls answered. COVID-19 vaccination hotline (number: 7077) also answered >97% of the calls on many days of the week during the study period. The percentage of calls handled by the COVID-19 vaccination hotline (number: 7077) improved to 62.17% by the date of writing this paper.

It is important to acknowledge that although this study provides valuable insights into the role of community call center during the COVID-19 pandemic in Qatar, the applicability of these findings to low- and middle-income countries (LMIC) may be subject to limitations. LMIC often face unique challenges, including resource constraints, technological disparities, and varied health care infrastructures. The digital divide prevalent in many LMIC could significantly exacerbate the access-to-technology gap, hindering the establishment and effectiveness of call centers. Moreover, the lack of access to robust health systems or timely responses in LMIC might impact the feasibility and sustainability of implementing similar call center solutions. The success of such initiatives in Qatar, being one of the richest countries globally, may not necessarily translate seamlessly to LMIC because of the nuanced economic, infrastructural, and health care disparities prevalent in those regions. Therefore, although call centers can be a valuable tool, particularly in health care emergencies, their feasibility and effectiveness must be carefully evaluated in the context of LMIC’s unique challenges and constraints.

Conclusions
This study highlights the significant role played by the community call center hotline (number: 16000) and the COVID-19 vaccination hotline (number: 7077) in effectively managing the challenges caused by the COVID-19 pandemic in Qatar. The community call center responded well to these challenges in Qatar by providing patients with an accessible, centralized resource as an alternative platform for telephone and video consultations and for managing a mass vaccination program. The findings demonstrate the call center’s operational efficiency to handle high call volumes and offer diverse consultations, effectively addressing a wide range of health concerns throughout different waves of the COVID-19 pandemic. Virtual consultations have emerged as a practical solution to empower patients with flexibility and convenience. Although video consultations experienced a temporary surge, their overall use gradually declined during the specified study period, reflecting patients’ preference for other consultation methods. Nevertheless, the community call center remained instrumental in managing various health care aspects, including chronic disease management and COVID-19–related queries. Moreover, the COVID-19 vaccination hotline played a major role in executing a successful mass immunization program. Prioritizing frontline health care workers and susceptible groups initially, the vaccination drive significantly contributed to reducing mortality rates in Qatar. This accomplishment can be attributed to the coordinated efforts of the community call center and other strategic measures implemented by the Ministry of Public Health of Qatar. As the COVID-19 pandemic recedes, it will be essential to adapt the community call center’s capacity and its services to address future health care challenges. Expanding community call center services to encompass specialized health services such as dermatology, counseling, health education, and psychological support could further enhance Qatar’s health care infrastructure. Our study emphasizes the critical significance of telemedicine and digital solutions in crisis response and health care delivery. The insights gained from the community call center in Qatar during the pandemic will provide valuable lessons for improving health care accessibility and emergency response strategies in the future. As health care systems transition to postpandemic norms, leveraging the knowledge and experience gained will foster resilience and efficacy in addressing upcoming challenges, ensuring the well-being of the population.
Acknowledgments
Publication fees was provided by Primary Health Care Corporation, Qatar. The authors received no specific funding for this work.

Data Availability
All data generated or analyzed during this study are included in this published paper.

Conflicts of Interest
None declared.

References


Abbreviations

GCC: Gulf Cooperation Council
LMIC: low- and middle-income countries
PHCC: Primary Health Care Corporation
QRCS: Qatar Red Crescent Society
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
Corrigenda and Addenda

Correction: Evaluating the Clinical Feasibility of an Artificial Intelligence–Powered, Web-Based Clinical Decision Support System for the Treatment of Depression in Adults: Longitudinal Feasibility Study

Christina Popescu1*, BSc, MSc; Grace Golden2*, David Benrimoh1*, MSc, CM, MD; Myriam Tanguay-Sela1, BASc; Dominique Slowey3, BSc; Eryn Lundrigan3, BSc; Jérôme Williams3, BSc; Bennet Desormeau1, BSc, MSc; Divyesh Kardani1, BSCE; Tamara Perez3, BSc, MSc; Colleen Rollins4, BSc; Sonia Israel1, BSc; Kelly Perlman1,3, BSc; Caitrin Armstrong1, BASc, MSc; Jacob Baxter3, BSc; Kate Whitmore3, BA; Marie-Jeanne Fradette3, BSc; Kaelan Fecarek-Hope5; Ghassen Soufi3, BSc; Robert Fratila1, BSc; Joseph Mehltretter3, MSc; Karl Looper3, MD; Warren Steiner3, MD; Soham Rej3, MSc, MD; Jordan F Karp5, MD; Katherine Heller6, BSc, MSc, PhD; Sagar V Parikh7, MD; Rebecca McGuire-Snieckus6, BA, MSc, PhD; Manuela Ferrari9, PhD; Howard Margolese3, MSc, CM, MD; Gustavo Turecki9, PhD, MD

1Aifred Health Inc., Montreal, QC, Canada
2University of Waterloo, Waterloo, ON, Canada
3McGill University, Montreal, QC, Canada
4University of Cambridge, London, United Kingdom
5University of Arizona, Tucson, AZ, United States
6Duke University, Durham, NC, United States
7University of Michigan, Ann Arbor, MI, United States
8Barts and the London School of Medicine, London, United Kingdom
9Douglas Mental Health University Institute, McGill University, Montreal, QC, Canada
*these authors contributed equally

Corresponding Author:
David Benrimoh, MSc, CM, MD
Aifred Health Inc.
1250 Rue Guy Suite #600
Montreal, QC, H3H 2T4
Canada
Phone: 1 5144637813
Email: david.benrimoh@mail.mcgill.com

Related Article:
Correction of: https://formative.jmir.org/2021/10/e31862
(JMIR Form Res 2024;8:e56570) doi:10.2196/56570

In “Evaluating the Clinical Feasibility of an Artificial Intelligence–Powered, Web-Based Clinical Decision Support System for the Treatment of Depression in Adults: Longitudinal Feasibility Study” (JMIR Form Res 2021;5(10):e31862), the authors noted one error.

They have been changed to read in the following manner:

The mean STAR-P and STAR-C total scores were listed as:

The mean STAR-P and STAR-C scores were 33.62 (SD 2.90) and 31.14 (SD 2.63) comparable with 38.4 (SD 12.0) and 31.5 (SD 6.9) in the original Scale to Assess Therapeutic Relationships in Community Mental Health Care study.

In “Assessing Physician and Patient Trust in the CDSS and Its Effect on the Clinician-Patient Relationship”, the STAR-P and STAR-C total scores were listed as:

The mean STAR-P and STAR-C scores were 33.62 (SD 2.90) and 31.14 (SD 2.63) comparable with 38.4 (SD 12.0) and 31.5 (SD 6.9) in the original Scale to Assess Therapeutic Relationships in Community Mental Health Care study.
The correction will appear in the online version of the paper on the JMIR Publications website on January 24, 2024 together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 19.01.24; this is a non–peer-reviewed article; accepted 19.01.24; published 24.01.24.

Please cite as:
Correction: Evaluating the Clinical Feasibility of an Artificial Intelligence–Powered, Web-Based Clinical Decision Support System for the Treatment of Depression in Adults: Longitudinal Feasibility Study
JMIR Form Res 2024;8:e56570
URL: https://formative.jmir.org/2024/1/e56570
doi: 10.2196/56570
PMID:
Web-Based Search Volume for HIV Tests and HIV-Testing Preferences During the COVID-19 Pandemic in Japan: Infodemiology Study

Rie Kanamori¹, MMedSc; Futaba Umemura², MMedSc; Kosuke Uemura¹, MD; Taiju Miyagami¹, MD, PhD; Simon Valenti¹, MD; Nobuyuki Fukui³, BA; Mayumi Yuda³, BE; Mizue Saita¹, MD, PhD; Hirotake Mori¹, MD, PhD; Toshio Naito¹, MD, PhD

¹Department of General Medicine, Faculty of Medicine, Juntendo University, Tokyo, Japan
²Department of Sports Medicine and Sportology, Graduate School of Medicine, Juntendo University, Tokyo, Japan
³Center for Promotion of Data Science, Graduate School of Medicine, Juntendo University, Tokyo, Japan

Corresponding Author:
Toshio Naito, MD, PhD
Department of General Medicine
Faculty of Medicine
Juntendo University
3-1-3 Hongo Bunkyo-ku
Tokyo, 113-8421
Japan
Phone: 81 3 5802 1190
Email: naito@juntendo.ac.jp

Abstract

Background: Research has found a COVID-19 pandemic–related impact on HIV medical services, including clinic visits, testing, and antiviral therapy initiation in countries including Japan. However, the change in trend for HIV/AIDS testing during the COVID-19 pandemic has not been explored extensively in the Japanese population.

Objective: This infodemiology study examines the web-based search interest for two types of HIV tests, self-test kits and facility-based tests, before and during the COVID-19 pandemic in Japan.

Methods: The monthly search volume of queried search terms was obtained from Yahoo! JAPAN. Search volumes for the following terms were collected from November 2017 to October 2018: “HIV test,” “HIV test kit,” and “HIV test health center.” The search term “Corona PCR” and the number of new COVID-19 cases by month were used as a control for the search trends. The number of new HIV cases in the corresponding study period was obtained from the AIDS Trend Committee Quarterly Report from the AIDS Prevention Foundation.

Results: Compared to the search volume of “corona-PCR,” which roughly fluctuated corresponding to the number of new COVID-19 cases in Japan, the search volume of “HIV test” was relatively stable from 2019 to 2022. When we further stratified by the type of HIV test, the respective web-based search interest in HIV self-testing and facility-based testing showed distinct patterns from 2018 to 2022. While the search volume of “HIV test kit” remained stable, that of “HIV test health center” displayed a decreasing trend starting in 2018 and has remained low since the beginning of the COVID-19 pandemic. Around 66%-71% of the search volume of “HIV test kits” was attributable to searches made by male internet users from 2018 to 2022, and the top three contributing age groups were those aged 30-39 (27%-32%), 20-29 (19%-32%), and 40-49 (19%-25%) years. On the other hand, the search volume of “HIV test health centers” by male users decreased from more than 500 from 2018 to 2019 to fewer than 300 from 2020 to 2022.

Conclusions: Our study found a notable decrease in the search volume of “HIV test health center” during the pandemic, while the search volume for HIV self-testing kits remained stable before and during the COVID-19 crisis in Japan. This suggests that the previously reported COVID-19–related decrease in the number of HIV tests mostly likely referred to facility-based testing. This sheds light on the change in HIV-testing preferences in Japan, calling for a more comprehensive application and regulatory acceptance of HIV self-instructed tests.
Introduction

HIV prevalence in Japan is <0.1% among adults aged 15-49 years, and the cumulative total of people diagnosed and living with HIV at the end of 2018 was approximately 30,000 [1]. A recent study using a national database found that 81.5% of people living with HIV who were on antiretroviral therapy (ART) had at least one chronic comorbidity [2]. According to recent national statistics, the proportion of people diagnosed and living with HIV on ART was 86% and 80% as of 2018 [3], lower than the first two 90-90-90 Joint United Nations Programme on HIV and AIDS (UNAIDS) goals, while the proportion of people living with HIV on ART who are virally suppressed was as high as 99% as of 2020 [4], reaching the UNAIDS goal. Evidence has suggested that earlier HIV diagnosis and linkage to treatment should be the core strategy in controlling the HIV epidemic [5]; however, the sudden and ongoing COVID-19 pandemic has overwhelmed the health care system globally and caused severe disruptions in medical service provision, ranging from elective procedures to routine surveillance for manageable diseases such as HIV infections [6-11]. Additionally, social distancing measures implemented early in the pandemic affected people’s access or willingness to visit health care facilities.

There are concerns that disruptions caused by the COVID-19 pandemic may lead to increased HIV incidence and mortality and pose challenges to the international community’s goal to eliminate the HIV or AIDS epidemic by 2030 [1]. Reports from Africa, Asia, Europe, Japan, and the United States have found a pandemic-related impact on HIV medical services, including clinic visits, testing, and antiviral therapy initiation [1-6]. In particular, the number of HIV tests conducted in health centers in Japan reduced from 142,000 in 2019 to 69,000 in 2020, with the difference amounting to a 50% decline in the first year of the COVID-19 pandemic [3]. The number of HIV consultations performed in public health centers also significantly declined in the second quarter of 2020 (32,565 tests) compared to the same period the year before (11,689 tests) [2]. According to recent national statistics, the proportion of people diagnosed and living with HIV at the end of 2018 was approximately 30,000 [1]. A recent study using a national database found that 81.5% of people living with HIV who were on antiretroviral therapy (ART) had at least one chronic comorbidity [2]. According to recent national statistics, the proportion of people diagnosed and living with HIV on ART was 86% and 80% as of 2018 [3], lower than the first two 90-90-90 Joint United Nations Programme on HIV and AIDS (UNAIDS) goals, while the proportion of people living with HIV on ART who are virally suppressed was as high as 99% as of 2020 [4], reaching the UNAIDS goal. Evidence has suggested that earlier HIV diagnosis and linkage to treatment should be the core strategy in controlling the HIV epidemic [5]; however, the sudden and ongoing COVID-19 pandemic has overwhelmed the health care system globally and caused severe disruptions in medical service provision, ranging from elective procedures to routine surveillance for manageable diseases such as HIV infections [6-11]. Additionally, social distancing measures implemented early in the pandemic affected people’s access or willingness to visit health care facilities.

In Japan, voluntary HIV testing is free of charge and anonymous, and offered as a package service bundled with pre- and posttest counseling at appointed public health facilities by law. Although the number of health center–conducted HIV tests increased steadily and peaked in 2008, a 30% decline was observed before the COVID-19 pandemic [8,12]. This is possibly due to the commercial availability of HIV self-tests in Japan, with the number of dry blood spot HIV self-tests increasing from 26,000 tests in 2005 to 91,000 tests in 2016 [7,12]. However, the HIV self-testing technique is not yet approved by the health authorities in Japan.

The internet is a common source of disease- and health-related information, and internet use influences care initiation and the treatment decisions of people living with HIV [13]. During the COVID-19 pandemic, infodemiology studies found that web-based search trends for symptoms associated with COVID-19 coincided with the disease outbreak [14]. Conversely, the global web-based interest in information on HIV/AIDS care services decreased [15]. However, the search engine trend for HIV/AIDS testing during the COVID-19 pandemic has not been explored extensively in the Japanese population. This study aims to characterize and compare the trend for web-based search interest in HIV testing before and during the pandemic to investigate whether there is a change in HIV/AIDS health-seeking behavior in Japan.

Data from the search engine provided by the Yahoo Japan Corporation will be used for this study because Yahoo! JAPAN is the most visited website in the country, and compared to most other countries, Japan uses Google less frequently.

Methods

Data Source

To examine the pattern of web-based search interests in HIV testing before and during the COVID-19 pandemic, the monthly search volume of selected search terms was determined based on the number of searches over a specified period extracted from Yahoo! JAPAN, which is one of the most used digital services in Japan, and compared to the rest of the world, Google is used less. According to the data, 68.9% of the population (aged >2 years) used Yahoo! JAPAN at least once a month between January and November 2021, while 65.1% used Google in the same period [16]. Search volume data were retrieved with authorized access from the Yahoo Japan Corporation server via the DS.INSIGHT People portal. It has been used in studies focusing on transition or trends in search behavior over time among different demographics (eg, gender, age, or prefecture) [17]. The user manual is available on the web [18].

The internet user population is the number of internet users throughout Japan, calculated based on the “Telecommunications Usage Trends Survey” published by the Ministry of Internal Affairs and Communications.

The quarterly number of new COVID-19 cases between January 2019 and December 2022 was obtained from an excerpt from the Japan Ministry of Health, Labour and Welfare [19].

The number of newly infected people with HIV between December 31, 2018, and June 26, 2022, was obtained from the AIDS Trend Committee Quarterly Report from the AIDS Prevention Foundation, API-Net AIDS Prevention Information Net [20].

Search Queries Used in the Analysis

The search term “HIV検査” (ie, “HIV test” in Japanese) was used to assess the web-based interest in general HIV testing from January 2020 to October 2021. The result is presented together with the search volume of “PCR” (ie,
“corona-PCR” in Japanese) and the number of new COVID-19 cases by month. The search volumes of “HIV検査キット” (ie, “HIV test kit” in Japanese) and “HIV検査 保健所” (ie, “HIV test health center” in Japanese) were compared to analyze the difference in web-based interest in self-instructed/postal and facility-based HIV testing from November 2018 to October 2022.

Data Standardization

The monthly number of searches per prefecture for each search query was obtained, adjusted by sex, and converted to standardized z scores according to the following formula:

\[ z = \frac{A - \mu}{\sigma} \]

“Query A” refers to the queried search term.

Statistical Analysis

Internet users were stratified according to age on the day of the search, sex, and year. Demographics were summarized as the number and percentage of patients for categorical variables. Descriptive statistics were summarized using Excel (Microsoft Corporation).

Ethical Considerations

This infodemiology study adhered to chapter 1, section 3, part 1, subsection (C), item 3 of the ethical guidelines for Medical and Health Research Involving Human Subjects of the Ministry of Health, Labour and Welfare of Japan [21]. In accordance with this guideline, since this study used previously anonymized and deidentified data, an ethical review was waived, and patient informed consent was not required.

Results

Between 2018 and 2022, 50% to 51% of overall adult internet users were male, and the user population was evenly distributed across different age groups by decade (11%-18% across different age groups in 2018, 12%-17% in 2019, 12%-18% in 2020, 13%-17% in 2021, and 13%-17% in 2022; Tables 1-3). The first nationwide state of emergency concerning COVID-19 was from April 7 to May 25, 2020, in Japan; since then, the increase in web-based interest in COVID-19 polymerase chain reaction testing (represented by the search volume of “corona-PCR”) roughly corresponded to the peaks and troughs of new COVID-19 infection cases in recent waves of the COVID-19 pandemic in Japan during mid 2021, early 2022, and late 2022. In contrast, the web-based interest in general HIV testing represented by the search term “HIV test” appeared to be relatively stable (Figure 1). Although the sex distribution of the “corona-PCR” search volume varied between 2020 and 2022, people aged 40-49 years consistently showed the highest web-based interest (26%-27%) for COVID-19 polymerase chain reaction testing, followed by those aged 50-59 years (21%-23%) and 30-39 years (15%-18%). In line with the characteristics of people with an increased risk of HIV infection, 57%-60% of the “HIV test” search volume between 2020 and 2022 were from male internet users, and people aged between 20 and 29 years (30%-31%) showed the most interest in general HIV testing within this period, followed by those aged between 30 and 39 years (24%-27%) and 40 and 49 years (17%-20%; Table 1).
Table 1. Search volumes of “HIV test” and “corona-PCR” by the sex and age of internet users in the years 2020, 2021, and 2022.

<table>
<thead>
<tr>
<th>Age group (year)</th>
<th>Year 2020, n (%)</th>
<th>Year 2021, n (%)</th>
<th>Year 2022 (January to October), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet usera</td>
<td>Search term “HIV検査”b</td>
<td>Search term “コロナpcr”c</td>
<td>Search term “HIV検査”</td>
</tr>
<tr>
<td>Overall</td>
<td>96,688,079 (100)</td>
<td>99,499,669 (100)</td>
<td>99,499,669 (100)</td>
</tr>
<tr>
<td>Search term “コロナpcr”c</td>
<td>2500 (100)</td>
<td>2700 (100)</td>
<td>2330 (100)</td>
</tr>
<tr>
<td>Search term “HIV検査”b</td>
<td>1300 (100)</td>
<td>1200 (100)</td>
<td>2380 (100)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>49,119,275 (51)</td>
<td>50,286,908 (51)</td>
<td>50,286,908 (51)</td>
</tr>
<tr>
<td>Female</td>
<td>47,568,804 (49)</td>
<td>49,212,761 (49)</td>
<td>49,212,761 (49)</td>
</tr>
<tr>
<td>Age group (year)</td>
<td>130 (10)</td>
<td>1200 (41)</td>
<td>1010 (43)</td>
</tr>
<tr>
<td>&lt;20</td>
<td>170 (7)</td>
<td>280 (10)</td>
<td>13,787,372 (14)</td>
</tr>
<tr>
<td>20-29</td>
<td>770 (30)</td>
<td>840 (31)</td>
<td>13,787,372 (14)</td>
</tr>
<tr>
<td>30-39</td>
<td>670 (27)</td>
<td>680 (25)</td>
<td>13,787,372 (14)</td>
</tr>
<tr>
<td>40-49</td>
<td>470 (19)</td>
<td>460 (17)</td>
<td>13,787,372 (14)</td>
</tr>
<tr>
<td>50-59</td>
<td>230 (9)</td>
<td>250 (9)</td>
<td>13,787,372 (14)</td>
</tr>
<tr>
<td>60-69</td>
<td>130 (5)</td>
<td>16,258,923 (16)</td>
<td>16,258,923 (16)</td>
</tr>
<tr>
<td>≥70</td>
<td>70 (3)</td>
<td>80 (3)</td>
<td>13,026,322 (13)</td>
</tr>
</tbody>
</table>

aThe internet user population is the number of internet users throughout Japan, calculated based on the “Telecommunications Usage Trends Survey” published by the Ministry of Internal Affairs and Communications [21].


dThe internet user population in 2022 uses the figures for 2021.
<table>
<thead>
<tr>
<th>Age group (years)</th>
<th>2018, n (%)</th>
<th>2019, n (%)</th>
<th>2020, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>95,590,219</td>
<td>107,480,629</td>
<td>96,688,079</td>
</tr>
<tr>
<td></td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>48,932,966</td>
<td>53,862,486</td>
<td>49,119,275</td>
</tr>
<tr>
<td></td>
<td>(51)</td>
<td>(50)</td>
<td>(51)</td>
</tr>
<tr>
<td>Female</td>
<td>46,657,253</td>
<td>53,618,143</td>
<td>47,568,804</td>
</tr>
<tr>
<td></td>
<td>(49)</td>
<td>(50)</td>
<td>(48)</td>
</tr>
<tr>
<td>Age group (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>12,760,379</td>
<td>13,710,441</td>
<td>13,386,518</td>
</tr>
<tr>
<td></td>
<td>(13)</td>
<td>(14)</td>
<td>(14)</td>
</tr>
<tr>
<td>20-29</td>
<td>12,381,162</td>
<td>12,514,936</td>
<td>11,777,729</td>
</tr>
<tr>
<td></td>
<td>(13)</td>
<td>(12)</td>
<td>(12)</td>
</tr>
<tr>
<td>30-39</td>
<td>14,321,086</td>
<td>14,167,245</td>
<td>13,547,255</td>
</tr>
<tr>
<td></td>
<td>(15)</td>
<td>(13)</td>
<td>(14)</td>
</tr>
<tr>
<td>40-49</td>
<td>18,127,686</td>
<td>18,214,081</td>
<td>17,429,796</td>
</tr>
<tr>
<td></td>
<td>(18)</td>
<td>(16)</td>
<td>(18)</td>
</tr>
<tr>
<td>50-59</td>
<td>14,891,256</td>
<td>15,903,690</td>
<td>15,451,255</td>
</tr>
<tr>
<td></td>
<td>(16)</td>
<td>(15)</td>
<td>(16)</td>
</tr>
<tr>
<td>60-69</td>
<td>12,991,502</td>
<td>14,688,962</td>
<td>12,716,666</td>
</tr>
<tr>
<td></td>
<td>(14)</td>
<td>(14)</td>
<td>(13)</td>
</tr>
<tr>
<td>≥70</td>
<td>10,117,148</td>
<td>18,281,274</td>
<td>12,378,860</td>
</tr>
<tr>
<td></td>
<td>(11)</td>
<td>(17)</td>
<td>(13)</td>
</tr>
</tbody>
</table>

aThe internet user population is the number of internet users throughout Japan, calculated based on the “Telecommunications Usage Trends Survey” published by the Ministry of Internal Affairs and Communications [22].
Table 3. Search volumes for “HIV test kit” and “HIV test health center” by the sex and age of the internet users by year (2021 and 2022).

<table>
<thead>
<tr>
<th>Year</th>
<th>Search term “hiv検査キット”</th>
<th>Search term “hiv検査保健所”</th>
<th>Year 2022 (January to October)</th>
<th>Search term “hiv検査キット”</th>
<th>Search term “hiv検査保健所”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>99,499,669 (100)</td>
<td>640 (100)</td>
<td>99,499,669 (100)</td>
<td>570 (100)</td>
<td>280 (100)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>50,286,908 (51)</td>
<td>290 (45)</td>
<td>50,286,908 (51)</td>
<td>390 (68)</td>
<td>100 (36)</td>
</tr>
<tr>
<td>Female</td>
<td>49,212,761 (49)</td>
<td>350 (55)</td>
<td>49,212,761 (49)</td>
<td>180 (32)</td>
<td>180 (64)</td>
</tr>
<tr>
<td>Age group (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>13,787,372 (14)</td>
<td>50 (8)</td>
<td>13,787,372 (14)</td>
<td>40 (7)</td>
<td>30 (11)</td>
</tr>
<tr>
<td>20-29</td>
<td>12,432,946 (13)</td>
<td>230 (35)</td>
<td>12,432,946 (13)</td>
<td>110 (19)</td>
<td>90 (32)</td>
</tr>
<tr>
<td>30-39</td>
<td>13,615,333 (14)</td>
<td>170 (27)</td>
<td>13,615,333 (14)</td>
<td>180 (32)</td>
<td>90 (32)</td>
</tr>
<tr>
<td>40-49</td>
<td>17,491,018 (17)</td>
<td>110 (17)</td>
<td>17,491,018 (17)</td>
<td>140 (25)</td>
<td>40 (14)</td>
</tr>
<tr>
<td>50-59</td>
<td>16,258,923 (16)</td>
<td>60 (9)</td>
<td>16,258,923 (16)</td>
<td>80 (14)</td>
<td>20 (7)</td>
</tr>
<tr>
<td>60-69</td>
<td>12,887,755 (13)</td>
<td>10 (2)</td>
<td>12,887,755 (13)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>≥70</td>
<td>13,026,322 (13)</td>
<td>10 (1)</td>
<td>13,026,322 (13)</td>
<td>20 (4)</td>
<td>10 (4)</td>
</tr>
</tbody>
</table>

*aThe internet user population is the number of internet users throughout Japan, calculated based on the “Telecommunications Usage Trends Survey” published by the Ministry of Internal Affairs and Communications [22].


dThe internet user population in 2022 uses the figures for 2021.

Figure 1. Monthly web-based search interest in HIV testing and COVID-19 genetic testing and the number of new COVID-19 cases from 2020 to 2022. The monthly search volumes of search terms “HIV検査” (HIV test in Japanese, blue line) and “コロナPCR” (corona-PCR in Japanese, red line) from January 2020 to October 2022 according to the search engine Yahoo! JAPAN are shown. The number of new COVID-19 cases in the corresponding months, according to the Ministry of Health, Labour and Welfare in Japan, is also shown in green bars in this figure.
Web-based interest in HIV testing was further refined as either facility-based HIV tests or self-instructed HIV-testing kits. The search volume for HIV facility-based testing was high in late 2018 and decreased in several instances over the next 3 years, especially in 2020 and 2022 (Figure 2). The search volume of “HIV test kit” (used as a representative search term for HIV self-testing) was lower than that for HIV facility-based testing at the end of 2018. However, the interest remained high over the next 4 years relative to facility-based testing, particularly during the COVID-19 pandemic (Figure 2). Around 66%-71% of the search volume of “HIV test kit” was attributable to searches made by male internet users between 2018 and 2022, and the top three contributing age groups were those 30-39 years (27%-32%), 20-29 years (19%-32%), and 40-49 years (19%-25%). The overall search volume of HIV test health centers decreased from 950 in 2018/2019 to 580 in 2020. The actual search volume for “HIV test kit” versus “HIV test health center” was 560 versus 280 among male internet users in 2020 and 590 versus 290 in 2021; while among female internet users the actual search volume for “HIV test kit” versus “HIV test health center” were similar: 230 versus 300 in 2020 and 310 versus 350 in 2021. Hence, the main population searching for “HIV test health center” shifted from male to female users (n=300, 52%) in 2020 and the following years (n=350, 55% in 2021 and n=180, 64% in 2022). Internet users aged 20-29 years (27%-35%), 30-39 years (24%-32%), and 40-49 years (14%-20%) were the top three representative age groups for the search volume of “HIV test health center” (Tables 2 and 3).

The quarterly number of newly infected HIV cases remained relatively stable since the declaration of a nationwide state of emergency concerning COVID-19, and the search interest remained high for self-testing relative to facility-based testing (Figure 3).

Figure 2. Monthly web-based search interest in HIV self-instructed and facility-based tests and the number of new COVID-19 cases from 2018 to 2022. The monthly search volumes of search terms “HIV検査キット” (HIV test kit in Japanese, orange line) and “HIV検査保健所” (HIV test health center in Japanese, purple line) from November 2018 to October 2022 according to the search engine Yahoo! JAPAN are shown. The number of new COVID-19 cases in the corresponding months, according to the Ministry of Health, Labour and Welfare in Japan, is shown with green bars.
Figure 3. Monthly web-based search interest in HIV self-instructed and facility-based tests and quarterly number of new infected HIV cases from 2019 to 2022. The monthly search volumes of search terms “HIV検査キット” (HIV test kit in Japanese, orange line) and “HIV検査 保健所” (HIV test health center in Japanese, purple line) from January 2019 to December 2022 according to the search engine provided by Yahoo! JAPAN are shown. The number of new HIV cases in the corresponding quarter, according to the AIDS Trend Committee Quarterly Report from the AIDS Prevention Foundation, API-Net AIDS Prevention Information Net, is shown with blue bars.

Discussion

The analysis in our study illustrated that the web-based search interest in general HIV testing remained stable in Japan during the ongoing pandemic. Furthermore, compared to a decrease in HIV facility-based testing, the web-based interest in HIV self-testing has not changed in Japan. The findings suggest that a proportion of the population interested in getting an HIV test or obtaining relevant information may have a lower preference for HIV facility-based testing. This could imply that individuals might have opted for HIV self-testing during the COVID-19 pandemic.

A prolonged decreasing trend in the search volume of “HIV test health center” from late 2018 was observed in this study. This change in search volume was related to the real-world situation reported by recent publications that underscored the situation of facility-based HIV testing during the pandemic [7,8,23]. A trend analysis study exploring data from 2015 to the second quarter of 2020 (during the nationwide state of emergency in Japan) reported a significant decline in the number of HIV tests performed by public health centers in the second quarter of 2020 (n=9584 vs n=35,908 in Q2 2019) [7] and that this decrease coincided with the increase in the number of new HIV cases with an AIDS diagnosis in Japan [7]. Similar situations were reported in other countries, where an approximately 31% to 50% reduction in the number of HIV tests performed by public health centers was observed between 2019 and 2020 due to difficulty accessing testing facilities, shortage of medical staff for HIV-testing services, or closure of HIV-testing facilities [23,24].

Our search trend showed that the web-based search volume for HIV self-tests was relatively unaffected in Japan. Furthermore, we noted the trend for HIV-testing preference was more evident among male internet users. In Japan, men who have sex with men (MSM) constitute more than 70% of newly diagnosed people living with HIV [3,25]; therefore, the difference in search volume for different HIV-testing services most likely reflects the changing preferences among male internet users. This is consistent with the real-world situation reported by other countries. Investigators focusing on the high-risk population’s (MSM) HIV-testing behavior in China during the COVID-19 pandemic found the use of HIV self-testing increased compared to facility-based HIV testing, which decreased by more than 50% overall [26,27]. Earlier investigations in Sweden and Japan showed the interest and demand for HIV self-testing in high-risk populations such as MSM even before the COVID-19 pandemic [12,28]. A study in France also reported that the number of HIV self-testing kits sold in 2021 increased by 3% compared to 2020 [29]. Together, these publications further emphasized the

In this study, interest in HIV self-testing has remained relatively high during the COVID-19 pandemic. Even though web-based search interest only indicates clinical information-seeking behavior, recent infodemiology studies, including that by Ornos et al [15], showed that search volume indices correlated positively with HIV prevalence and negatively with financial and health care service status. Another recent study in Japan by Ishimaru et al [30] also observed a positive correlation between internet search frequency for HIV/AIDS–related terms and the number of voluntary tests.

The change in search volume on HIV-testing preferences in Japan, which highlights the availability, accessibility, and regulatory approval status of HIV self-tests, may be imperative to reduce new HIV infection [7,8]. The pros and cons of alternative HIV-testing approaches such as the use of dried blood spot (DBS) test cards delivered by postal service have been investigated in Japan [8]. So far, DBS-based tests have not been approved as clinical samples for HIV testing in Japan; however, their use has steadily increased in the last two decades, most likely due to convenient and easy self-preparation of blood spots without the need to visit medical facilities [8]. Although the use of DBS is less sensitive than with a plasma sample, the feasibility and reliability of postal DBS have been preliminarily demonstrated in an outreach study in Japan [12]. Other HIV testing to be considered should include the introduction of HIV self-testing, DBS, or oral swabs as an alternative to plasma or serum specimens. However, since HIV self-testing is not yet approved by health authorities, individuals with positive test results may not have been referred to medical facilities for consultations after the test or received appropriate HIV care and treatment. This concern was illustrated in the study by Ejima et al [7], which reported that both the number of HIV tests and consultations performed by public health centers declined during the COVID-19 pandemic in Japan. We believe relevant supporting services, for instance, the provision of web-based counseling services and referral services after HIV self-testing, would be necessary to mitigate the potential concerns and negative impact of using HIV test kits without counseling and medical follow-up. Soon, the development of artificial intelligence chatbots may be used to provide real-time instruction and counseling for HIV self-testing users, which may offer potential solutions to this problem [31].

Several factors may limit the generalizability of the preliminary findings presented in this report. First, the data source was solely from Yahoo! JAPAN, and the internet searches conducted using other search engines were not included. Second, the data obtained from a web-based search engine may be subject to the nonrepresentative sampling or methodology bias inherent to the search platform. Third, there could be considerable differences in clinical information-seeking behavior between the internet user population and actual men and women who are affected by HIV; direct and causal relationships cannot be inferred from this study. Nonetheless, most previous studies using search engines have used Google Trends data, but in Japan, Yahoo! JAPAN is the most visited digital service in the country. In addition, sex and prefecture adjustment is possible with Yahoo! JAPAN data, making it appropriate to research topics likely affected by sex differences. Therefore, our preliminary infodemiology using the search engine provided by Yahoo! Japan Corporation is likely representative of the general trend in Japan.

Our infodemiology study indicated that there was a notable decrease in search volume for HIV facility-based testing during the COVID-19 pandemic. Further, a change in HIV-testing preference and interest in HIV self-testing was noted in Japan. To fully delineate and comprehend the changes in HIV-testing behavior, the situation should be continuously monitored and validated by clinical studies.

Acknowledgments
The authors would like to thank 4DIN Ltd, Japan, for the data analysis and for assisting with preparing the manuscript and other materials for journal submission. This work was supported by the Ministry of Health, Labour and Welfare Research on HIV/AIDS (program grant JPMH21HB1005).

Data Availability
The data sets generated or analyzed during this study are not publicly available, complying with the Yahoo! Japan Corporation regulation, but they are available from the corresponding author upon reasonable request.

Authors’ Contributions
FU and TN conceptualized this study. RK and KU are responsible for the methodology. NF and MY collected the data. RK, FU, KU, TM, SV, MS, and HM analyzed the data. RK, NF, MY, and TN prepared the initial draft. RK, FU, KU, TM, SV, MS, HM, and TN critically revised the manuscript. HM and TN obtained the funding. MS was responsible for project administration. All authors approved this version of the manuscript to be submitted.

Conflicts of Interest
None declared.

References
[30] Ishimaru et al [30] also observed a positive correlation between internet search frequency for HIV/AIDS–related terms and the number of voluntary tests.


Abbreviations

- ART: antiretroviral therapy
- DBS: dried blood spot
- MSM: men who have sex with men
- UNAIDS: Joint United Nations Programme on HIV and AIDS