Contents

Original Papers

An mHealth Platform for Augmenting Behavioral Health in Primary Care: Longitudinal Feasibility Study (e36021)
Khatiya Moon, Michael Sobolev, Megan Grella, George Alvarado, Manish Sapra, Trever Ball ................................. 4

Exploring a Need for a Cardiometabolic Disease Staging System as a Computerized Clinical Decision Support Tool: Qualitative Study (e37456)
Aizhan Karabukayeva, Jami Anderson, Allyson Hall, Sue Feldman, Tapan Mehta ........................................................ 12

Improvements in Depression Outcomes Following a Digital Cognitive Behavioral Therapy Intervention in a Polychronic Population: Retrospective Study (e38005)
Aarathi Venkatesan, Benjamin Forster, Prasanna Rao, Melissa Miller, Michael Scahill .............................................. 23

Optimization of the Chronic Kidney Disease–Peritoneal Dialysis App to Improve Care for Patients on Peritoneal Dialysis in Northeast Thailand: User-Centered Design Study (e37291)
Eakalak Lukkanalikitkul, Sawinee Kongpetch, Wijittra Chotmongkol, Michael Morley, Sirirat Anutrakulchai, Chavis Srichan, Bandit Thinkhamrop, Theenatchar Chungthom, Pongsai Wiangnon, Wilaiphorn Thinkhamrop, Katharine Morley ........................................... 37

Experimental Implementation of NSER Mobile App for Efficient Real-Time Sharing of Prehospital Patient Information With Emergency Departments: Interrupted Time-Series Analysis (e37301)
Kiyomitsu Fukaguchi, Tadahiro Goto, Tadatsugu Yamamoto, Hiroshi Yamagami ......................................................... 51

Chinese Version of the Mobile Health App Usability Questionnaire: Translation, Adaptation, and Validation Study (e37933)
Yi Shan, Meng Ji, Wenxiu Xie, Rongying Li, Xiaobo Qian, Xiaomin Zhang, Tianyong Hao ............................................ 62

Perspectives of Patients and Therapists on Social Media and Digital Data Use in Mental Health Therapy: Thematic Analysis (e32103)
Lauren Southwick, Rebecca Suh, Elissa Kranzler, Megan Bradley, Raina Merchant ................................................... 75

A German Smartphone-Based Self-management Tool for Psoriasis: Community-Driven Development and Evaluation of Quality-of-Life Effects (e32593)
Lea Brandl, Claudia Liebram, Wendelin Schramm, Monika Pobiruchin ................................................................. 87

Exploring Users’ Experiences With a Quick-Response Chatbot Within a Popular Smoking Cessation Smartphone App: Semistructured Interview Study (e36869)
Alice Alphonse, Kezia Stewart, Jamie Brown, Olga Perski ................................................................. 98
Smartphone Ownership, Smartphone Utilization, and Interest in Using Mental Health Apps to Address Substance Use Disorders: Literature Review and Cross-sectional Survey Study Across Two Sites (e38684)
Michael Hsu, Bianca Martin, Saeed Ahmed, John Torous, Joji Suzuki. 111

Feasibility of Mobile Health and Social Media–Based Interventions for Young Adults With Early Psychosis and Clinical Risk for Psychosis: Survey Study (e30230)
Olivia Franco, Monica Calkins, Salvatore Giorgi, Lyle Ungar, Raquel Gur, Christian Kohler, Sunny Tang. 123

Development of a Maternal and Child mHealth Intervention With Aboriginal and Torres Strait Islander Mothers: Co-design Approach (e33541)
Sarah Perkes, Belinda Huntriss, Noelene Skinner, Bernise Leece, Rosie Dobson, Joerg Mattes, Kerry Hall, Billie Bonevski. 130

Developing Educational Animations on HIV Pre-exposure Prophylaxis (PrEP) for Women: Qualitative Study (e33978)
Anna Young, Timothee Fruhauf, Obianuju Okonkwo, Erin Gingher, Jenell Coleman. 147

Adaptation of a Problem-solving Program (Friendship Bench) to Treat Common Mental Disorders Among People Living With HIV and AIDS and on Methadone Maintenance Treatment in Vietnam: Formative Study (e37211)
Ha Tran, Ha Nong, Thuy Tran, Teresa Filipowicz, Kelsey Landrum, Brian Pence, Giang Le, Minh Nguyen, Dixon Chibanda, Ruth Verhey, Vivian Go, Hien Ho, Bradley Gaynes. 154

Preliminary Real-World Evidence Supporting the Efficacy of a Remote Neurofeedback System in Improving Mental Health: Retrospective Single-Group Pretest-Posttest Study (e35636)
Jocelyne Whitehead, Ron Neeman, Glen Doniger. 168

Novel Implementation Strategy to Electronically Screen and Signpost Patients to Health Behavior Apps: Mixed Methods Implementation Study (OptiMine Study) (e34271)
Zarnie Khadjesari, Tracey Brown, Alex Ramsey, Henry Goodfellow, Sherine El-Toukhy, Lorien Abroms, Helena Jopling, Arden Dierker Viik, Michael Amato. 182

A Web Platform for Standardized Data Acquisition, Processing, and Export in the Child Psychopathology Clinical Routine (MedicalBIT): Design and Implementation Study (e36757)
Paola Colombo, Silvia Busti Ceccarelli, Stefano Pacchiarini, Stefano Cribellati, Massimo Molteni. 195

Participant Experiences of a COVID-19 Virtual Clinical Study Using the Current Health Remote Monitoring Platform: Case Study and Qualitative Analysis (e37567)
Juliana Pugmire, Jessie Lever Taylor, Matt Wilkes, Adam Wolfberg, Nicole Zahradka. 207

Measures of Daily Activities Associated With Mental Health (Things You Do Questionnaire): Development of a Preliminary Psychometric Study and Replication Study (e38837)
Nickolai Titov, Blake Dear, Madelyne Bisby, Olav Nielssen, Lauren Staples, Rony Kayrouz, Shane Cross, Eyal Karin. 222

Standardizing Primary Health Care Referral Data Sets in Nigeria: Practitioners’ Survey, Form Reviews, and Profiling of Fast Healthcare Interoperability Resources (FHIR) (e28510)
Emeka Chukwu, Lalit Garg, Nkiruaka Obande-Ogbuinya, Vijay Chattu. 238

Understanding the Security and Privacy Concerns About the Use of Identifiable Health Data in the Context of the COVID-19 Pandemic: Survey Study of Public Attitudes Toward COVID-19 and Data-Sharing (e29337)
Charlotte Summers, Frances Griffiths, Jonathan Cave, Arjun Panesar. 251

Development of a Quality Management Model and Self-assessment Questionnaire for Hybrid Health Care: Concept Mapping Study (e38683)
Rosian Tossaint-Schoenmakers, Marise Kasteleyn, Anneloek Rauwerdink, Niels Chavannes, Sofie Willems, Esther Talboom-Kamp. 261
Assessing the Views and Needs of People at High Risk of Gestational Diabetes Mellitus for the Development of Mobile Health Apps: Descriptive Qualitative Study (e36392)
Beibei Duan, Zhe Liu, Weiwei Liu, Baohua Gou ................................................................. 279

Health-Related Quality of Life Among Members Using an On Demand Behavioral Health Platform: Pilot Observational Study (e35352)
Emily Shih, Brandon Aylward, Sarah Kunkle, Grant Graziani .............................................. 291

Comparison of the Impact of Insulin Degludec U100 and Insulin Glargine U300 on Glycemic Variability and Oxidative Stress in Insulin-Naive Patients With Type 2 Diabetes Mellitus: Pilot Study for a Randomized Trial (e35655)
Pavle Vrebalov Cindro, Mladen Krnić, Darko Modun, Jonatan Vuković, Tina Tišinović Kurir, Goran Kardum, Doris Rušić, Ana Šešelja Perišin, Josipa Bukić .......................................................... 300

The Association Between COVID-19 Information Sources and Stigma Against Health Care Workers Among College Students: Cross-sectional, Observational Study (e35806)
Miharu Nakanishi, Mai Sakai, Gen Takagi, Keita Toshi, Koubun Wakashima, Hatsumi Yoshii .......................................................... 310

Exploring Public Perceptions of Dental Care Affordability in the United States: Mixed Method Analysis via Twitter (e36315)
Shahen Yashpal, Ananditha Raghunath, Nihan Gencerlıler, Lorel Burns .................................................. 321

Corrigenda and Addenda

Correction: The Associations Between Racially/Ethnically Stratified COVID-19 Tweets and COVID-19 Cases and Deaths: Cross-sectional Study (e40365)
Xiaohui Liu, Bandana Kar, Francisco Montiel Ishino, Tracy Onega, Faustine Williams .................................................. 319
An mHealth Platform for Augmenting Behavioral Health in Primary Care: Longitudinal Feasibility Study

Khatiya Chelidze Moon1, MD; Michael Sobolev2, PhD; Megan Grella3, LCSW, MBA; George Alvarado3, MD; Manish Sapra3, MD; Trever Ball3, PhD

1Zucker Hillside Hospital, Glen Oaks, NY, United States
2Cornell Tech, New York, NY, United States
3Northwell Health, Manhasset, NY, United States

Corresponding Author:
Khatiya Chelidze Moon, MD
Zucker Hillside Hospital
Kaufmann Building, Suite k204
75-59 263rd Street
Glen Oaks, NY, 11004
United States
Phone: 1 7184704367
Email: kmoon2@northwell.edu

Abstract

Background: The collaborative care model is a well-established system of behavioral health care within primary care settings. There is potential for mobile health (mHealth) technology to augment collaborative behavioral health care in primary care settings, thereby improving scalability, efficiency, and clinical outcomes.

Objective: We aimed to assess the feasibility of engaging with and the preliminary clinical outcomes of an mHealth platform that was used to augment an existing collaborative care program in primary care settings.

Methods: We performed a longitudinal, single-arm feasibility study of an mHealth platform that was used to augment collaborative care. A total of 3 behavioral health care managers, who were responsible for coordinating disease management in 6 primary care practices, encouraged participants to use a mobile app to augment the collaborative model of behavioral health care. The mHealth platform’s functions included asynchronous chats with the behavioral health care managers, depression self-report assessments, and psychoeducational content. The primary outcome was the feasibility of engagement, which was based on the number and type of participant-generated actions that were completed in the app. The primary clinical end point was a comparison of the baseline and final assessments of the Patient Health Questionnaire-9.

Results: Of the 245 individuals who were referred by their primary care provider for behavioral health services, 89 (36.3%) consented to app-augmented behavioral health care. Only 12% (11/89) never engaged with the app during the study period. Across all participants, we observed a median engagement of 7 (IQR 12; mean 10.4; range 0-130) actions in the app (participants: n=78). The chat function was the most popular, followed by psychoeducational content and assessments. The subgroup analysis revealed no significant differences in app usage by age ($P=.42$) or sex ($P=.84$). The clinical improvement rate in our sample was 73% (32/44), although follow-up assessments were only available for 49% (44/89) of participants.

Conclusions: Our preliminary findings indicate the moderate feasibility of using mHealth technology to augment behavioral health care in primary care settings. The results of this study are applicable to improving the design and implementation of mobile apps in collaborative care.

(JMIR Form Res 2022;6(7):e36021) doi:10.2196/36021

KEYWORDS

collaborative care; mobile health; psychiatry; depression; virtual care; psychoeducation; mobile app; mobile phone
Introduction

Background

The reach of behavioral health services is insufficient for meeting the needs of the population [1,2]. The collaborative care model (CoCM) is a framework that attempts to meet this vast need for behavioral health services by embedding these services in primary care settings [3]. The CoCM is a system of outcome-driven, stepwise care for systematically identifying individuals who would benefit from behavioral health treatment and supporting primary care clinicians in their management. The model has been adapted by many health systems since its introduction in the 1990s and is considered best practice [4-6]. Unfortunately, there are challenges that limit the scalability of the CoCM, including financial and operational barriers [7,8]. Innovative approaches are needed to support these existing models of behavioral health care [9].

There is emerging evidence that digital and mobile health (mHealth) technologies have the potential to improve the reach of the CoCM [9-12]. Given the near ubiquity of smartphones with app capabilities, health systems are increasingly interested in understanding whether these tools can be harnessed to further extend collaborative care [13,14]. There are several meaningful ways that mobile apps could be used to augment collaborative care. The CoCM relies strongly on a measurement-based system of care in which validated clinical assessments (e.g., the Patient Health Questionnaire-9 [PHQ-9]) are regularly collected to assess clinical responses [15]. mHealth platforms could help decrease care providers’ workload by automating the collection of these measures [12,13]. Apps could also support clinical decision-making by collecting clinical information more frequently than what is currently possible [10]. Moreover, apps could facilitate more frequent communication between patients and care providers [16,17]. Finally, apps can act as repositories for educational materials and self-guided modules for reinforcing concepts that are learned in therapy and promoting patient engagement [13,18-20]. All of these factors have the potential to increase patient engagement in care and, eventually, result in improved clinical outcomes.

Despite the theoretical benefits of app-augmented collaborative care, relatively little is still known about the feasibility of app usage in collaborative care. It is unknown whether patients in collaborative care settings are likely to use apps, what features of mobile apps are the most beneficial in this setting, and what patient population(s) may be the most likely to benefit from app-augmented collaborative care [11]. For example, in the broader literature on mHealth, there is concern that older individuals may be less familiar with technology and may therefore be less inclined to engage with it [21]; however, this has not been systematically examined, to our knowledge, in the collaborative care setting. The limited existing studies suggest that overall, app usage among patients in collaborative care can be variable [9,12]. Understanding app usage is applicable to the optimization of mHealth platform interventions and their implementation in collaborative care [22]. If known, this information could lay the groundwork for improving the design and implementation of mobile apps in collaborative care.

Objectives

This study describes a feasibility study of the Valera Health mobile platform and app (Valera Health Inc), which was used to augment collaborative care within primary care practices in a large health care system. Our primary aim was to assess the feasibility of app usage, which was measured based on engagement. A secondary outcome was preliminary clinical improvement in depression scores.

Methods

Ethics Approval

The study methods were approved by the Institutional Review Board of Northwell Health (approval number: 20-0545-NH). Informed consent was not sought due to the retrospective nature of the study.

Study Overview

This was a retrospective review of a longitudinal, single-arm implementation initiative wherein individuals who were referred to the collaborative care program by their primary care providers (PCPs) were invited to participate in app-augmented collaborative care by the behavioral health care manager (BHCM). Individuals who were qualified to participate and agreed to do so were asked to download the Valera Health mobile app. Participants were told to use the app to complete in-app PHQ-9 measures, which were sent by the BHCM at preset monthly intervals; communicate with the BHCM through asynchronous chats as needed; and access the psychoeducational content in the app. Participants also experienced all usual collaborative care interventions, as described in the Study Setting section, including office visits and telephone contacts with the BHCM, short-term psychotherapy, care coordination, psychiatric case reviews, follow-ups with their PCPs as indicated, and the prescription of recommended psychiatric medications if indicated. PHQ-9 assessments were able to be completed through the app or on paper during office visits with the BHCM. For our primary outcome—feasibility—usage data on the number of user actions that were completed in the app were recorded by the app throughout the participation period. For our secondary outcome—clinical improvement—baseline and final PHQ-9 scores were compared.

Study Setting

The study was conducted in a large, primarily suburban, academic health care system with multiple affiliated primary care practices. The primary care practices that were involved in this study provide behavioral health services through a system that was modeled after the CoCM introduced in the Improving Mood—Providing Access to Collaborative Treatment trial [6]. Briefly, in this model, patients presenting for routine primary care are systematically screened for depression by their PCPs to augment collaborative care within primary care practices in the BHCM who is physically embedded in the clinic. The BHCM maintains a registry of patients, tracks outcomes via serial PHQ-9 assessments, provides time-limited psychotherapy,
coordinates referrals to continued treatment and/or a higher level of care when necessary, and liaises with a psychiatrist who provides remote supervision to multiple BHCMs. Psychopharmacologic recommendations are relayed to the PCP, who remains the prescriber and clinician of record.

Recruitment

Recruitment was planned in 2 phases. The first phase lasted from November 2018 to June 2019 and included 1 primary care practice with 1 BHCM. A total of 5 additional practices and 2 BHCMs were added in phase 2, which lasted from November 2019 to March 2020. There were no differences in procedures between the two phases except for the number of clinics and BHCMs involved. Patients who were referred to the collaborative care program were invited to participate in app-augmented collaborative care by the BHCM during initial appointments. During recruitment at the initial visits, the BHCM guided the participants through the process of downloading and using the app, answered any initial questions, and provided written instructions on the use of the app. After the initial visits, the BHCM was available by phone to troubleshoot the app as needed. Recruitment was halted in March 2020 when the social distancing measures that were required to prevent the spread of COVID-19 in New York resulted in the remote provision of ambulatory behavioral health services.

Individuals who declined to participate or were excluded received usual collaborative care, as described above, with the Valera Health mobile app as augmentation.

Individuals were included if they were adults with a diagnosis of depression or anxiety. Individuals were also excluded if they did not speak English or had severe mental illnesses, suicidal or violent ideation, or substance abuse disorders. Also excluded were children and individuals who required a referral to a higher level of care or continued treatment after the completion of the program.

Participation flow is illustrated by Figure 1. Between November 2019 and March 2020, a total of 245 individuals were referred by their PCPs to the collaborative care program for behavioral health services. Further, 58% (n=142) of these patients were eligible for and were recruited to participate in our study using the Valera Health mobile app (Figure 2), and 62.7% (89/142) of recruited patients consented to participate; 34 consented during the first phase of piloting the Valera Health mobile app, and 55 consented during the second phase. The time required to train patients in the use of the app was a barrier to recruitment among a sizable minority of individuals in the target population (38/245, 15.5%; Figure 1). In addition, a portion of our eligible patient population was unable to participate due to technical barriers (22/142,15.5%; Figure 1). Further, 7 individuals declined to participate due to privacy concerns (Figure 1). Participants were mostly female (60/89, 67%) and middle-aged (mean 38.6, SD 14 years). Participants were enrolled in the study for an average of 22 weeks.

Figure 1. The flow of patient participation in the Valera Health mobile app pilot for behavioral health.
Intervention
The Valera Health mobile app is an English-language secure platform with several functionalities. Figure 2 shows screenshots of the app. First, the app automatically sends PHQ-9 assessments to participants at monthly intervals that are preset by the BHCM. Second, the app allows for secure asynchronous messaging between participants and the BHCM. BHCMs typically responded to chat messages from participants within 1 business day. The app also contains psychoeducational content. This content includes written material about common behavioral health conditions such as depression and anxiety, education about treatments like medication and psychotherapy, instructional guides on topics such as mindfulness, and video and audio clips on these topics. The BHCM had the option of prompting the participants to access psychoeducational content that was relevant to the participants’ care via the app. The Valera Health app was not integrated into the electronic health record (EHR). Relevant clinical information from the app, such as PHQ-9 scores, was documented into the EHR by the BHCM.

Outcomes
We assessed the feasibility of the mHealth intervention by investigating engagement with the app and improvement in clinical outcomes. The primary outcome—the feasibility of engagement—was assessed based on app usage, which was measured as the number of participant-generated actions completed in the app. Possible participant-generated actions were (1) the in-app completion of PHQ-9 assessments, (2) the sending of a chat message, or (3) the accessing of in-app psychoeducational material. App usage was monitored throughout the study by the Valera Health app. Engagement was stratified by age and sex. A secondary outcome was clinical improvement, which we defined as a final PHQ-9 score of less than 10 or a greater than 50% reduction in PHQ-9 scores. Baseline PHQ-9 scores were assessed during intake by the BHCM. The final PHQ-9 scores were the last ones recorded for the participants and were extracted from the EHR. Any PHQ-9 assessment, whether it was completed through the app or on paper during office visits with the BHCM, was considered in the analysis of clinical improvement.

Data Analysis
Data were analyzed by using Microsoft Excel, and an $\alpha$ of .05 was set as the a priori level of significance. Simple descriptive statistics of app usage were used to report on feasibility and engagement metrics. A Kruskal-Wallis equality-of-populations rank test of median differences in total mobile app actions was used to analyze differences in app usage between age and sex groups. For our secondary outcome—clinical improvement—a 2-tailed, 2-sample Welch $t$ test with unequal variances was used to analyze the difference between baseline and follow-up PHQ-9 scores.

Results
Table 1 describes the brief demographic and engagement characteristics of participants. Only 12% (11/89) never engaged with the app during the trial. Across all participants, we observed a median engagement of 7 (IQR 12; mean 10.4; range 0-130) actions in the app. At least 1 action in the mobile app was completed by 87% (78/89) of participants. Psychoeducational content was reviewed by 75% (67/89) of participants (number of articles reviewed: median 2, IQR 4; range 0-18). Chat messages were sent by 62% (55/89) of participants. Participants sent a median of 1 (IQR 5; range 0-115) chat message used for either scheduling an appointment or reporting symptoms. A baseline PHQ-9 score was reported for 97% (86/89) of study participants. However, follow-up PHQ-9 assessments were
available only for 49% (44/89) of participants. PHQ-9 scores improved from baseline to follow-up for 73% (32/44) of participants for whom we had baseline and follow-up PHQ-9 scores (n=44). The percentages of participants with improved PHQ-9 scores were not different by sex (P=.53) or age groups (18-35, 36-55, and ≥56 years: P=.90), although as previously noted, the sample of participants with recorded follow-up PHQ-9 scores was notably smaller than the sample of participants with baseline PHQ-9 scores.

Table 1. Demographic and engagement characteristics of behavioral health patients who participated in the Valera Health mobile app study (N=89).

<table>
<thead>
<tr>
<th>Gender</th>
<th>Value, n (%)</th>
<th>Age (years), mean (SD)</th>
<th>Baseline PHQ-9 score, mean (SD)</th>
<th>Follow-up PHQ-9 score, mean (SD)</th>
<th>P value</th>
<th>Total mobile app actions, median (IQR)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>89 (100)</td>
<td>38.6 (14.5)</td>
<td>11.5 (5.3)</td>
<td>8.6 (4.6)</td>
<td>7 (12)</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>60 (67)</td>
<td>38.4 (14.6)</td>
<td>11.5 (5.3)</td>
<td>9.0 (4.8)</td>
<td>6.5 (12)</td>
<td>.84</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>29 (33)</td>
<td>39.1 (14.6)</td>
<td>11.7 (5.5)</td>
<td>7.5 (3.9)</td>
<td>8 (11)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>18-35</td>
<td>44 (51\textsuperscript{a})</td>
<td>N/A</td>
<td>11.8 (5.1)</td>
<td>8.7 (4.0)</td>
<td>8 (14)</td>
<td>.42</td>
</tr>
<tr>
<td></td>
<td>36-55</td>
<td>31 (36\textsuperscript{a})</td>
<td>N/A</td>
<td>11.1 (4.9)</td>
<td>8.3 (5.9)</td>
<td>6 (9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt;55</td>
<td>12 (14\textsuperscript{a})</td>
<td>N/A</td>
<td>11.7 (7.2)</td>
<td>8.9 (4.3)</td>
<td>5.5 (11.5)</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}PHQ-9: Patient Health Questionnaire-9.
\textsuperscript{b}n=86.
\textsuperscript{c}n=44.
\textsuperscript{d}A 2-sample Welch t test with unequal variances between baseline and follow-up PHQ-9 scores.
\textsuperscript{e}A Kruskal-Wallis equality-of-populations rank test of median differences in the total mobile app actions.
\textsuperscript{f}N/A: not applicable.
\textsuperscript{g}The denominator for this percentage is 87.

Discussion

Principal Findings

This study reports on the evaluation of one of the first implementations of mobile app–augmented care within a collaborative care program. The overarching purpose of this feasibility study was to understand whether patients in collaborative care are likely to participate in app-augmented care, what features of the app are used, and whether demographic differences exist among app users in this context. Overall, our results indicate the acceptability and feasibility of app usage; the overwhelming majority of participants (78/89, 87%) used the app at least once, and a modest median of 7 actions were completed in the app. Encouragingly, all of the features of the app were used at similar rates. In particular, the psychoeducational materials and chat feature were both popular functions, suggesting that the app may indeed act to reinforce the concepts that are learned during clinical encounters and can enhance communication between patients and care providers as postulated in the literature [13]. There was no significant difference in app usage by age group (P=.42). Finally, we were encouraged by the finding that privacy—a concern in the broader literature surrounding app usage—was an infrequent cause for declining to participate (n=7) in our study, further suggesting acceptability [24,25].

With regard to our secondary outcome, the preliminary clinical outcomes of app-augmented care in this study were encouraging, with 73% (32/44) of participants for whom follow-up PHQ-9 data were available experiencing improvements (n=44). Our findings on clinical outcomes are limited by the considerable drop-off in the number of participants who completed a follow-up PHQ-9 assessment. The reasons for such decreases are not completely known and include app attrition, which is consistent with previous literature showing that attrition is a challenge to the implementation of mobile technologies [12,26]. The missing data also introduced bias into the study, as the population of participants who did not complete follow-up assessments may not be random.

Implications

Our findings reveal several issues that deserve consideration and optimization prior to subsequent implementation efforts. First, despite the prevalence of smartphones, a nontrivial portion of our eligible patient population was unable to participate due to technical barriers (22/142, 15.5%). In addition, the time required to train patients in the use of the app was a barrier to recruitment among a sizable minority of individuals in the target population (38/245, 15.5%). To address these issues, previous literature has postulated the need for a digital health navigator—a new team member with expertise in digital and mobile strategies who can help educate patients on the use of these tools, thereby reducing the burden among staff who may lack this expertise and have insufficient time to address these topics during appointments [14,27,28]. Digital health navigators can be instrumental to training staff in rapidly developing competencies for mHealth [29]. The CoCM, which already operates in a framework of interdisciplinary collaboration, may
be uniquely suited to the adoption of the digital health navigator role in the future.

Second, future implementation efforts should carefully consider measures for mitigating the potential unintended negative consequences of app use. For example, the chat function could give participants the perception of continuous access to the clinician. Though most participants in our study used the chat a moderate amount of times, there was substantial variability, with some participants sending more than 100 messages. Care provider burnout is a concern in the face of such significant increases in patient communication. The future implementation of this and similar technologies would benefit from integration with existing health information systems and care provider workflows to better support care provider decision-making.

Finally, while our results indicate preliminary evidence that patients are willing to participate in app-augmented collaborative care, there exist many opportunities for the optimization of engagement. For example, our intervention allowed the BHCM to nudge patients to engage with psychoeducational materials. Higher engagement can potentially be achieved by personalizing technology to deliver the right content in the right amount and at the right time [11,30,31]. Future research should identify pathways for personalization and investigate its effects on engagement and clinical improvement.

**Limitations**

Our study has several limitations. First, qualitative data for contextualizing findings were not systematically collected as a part of this study. Second, the experiences of the BHCMs were also not systematically evaluated. This study examined app usage among participants, which was stratified by age and sex; however, other demographic factors of potential interest, such as race, ethnicity, insurance status, were not tracked. Finally, this was a feasibility study; thus, conclusions are limited by the lack of a control group and the unknown characteristics of nonparticipants. Future research should employ a rigorous clinical trial involving patient and/or clinic randomization to evaluate clinical effectiveness.

**Conclusions**

In conclusion, our preliminary findings indicate the moderate feasibility of using mHealth technology to augment behavioral health care in primary care settings. The results of this study are applicable to improving the design and implementation of mobile apps in collaborative care.

**Acknowledgments**

This pilot study was sponsored by a Delivery System Reform Incentive Payment grant from New York State.

**Conflicts of Interest**

None declared.

**References**


Abbreviations

BHCM: behavioral health care manager
CoCM: collaborative care model
EHR: electronic health record
mHealth: mobile health
PCP: primary care provider
PHQ-9: Patient Health Questionnaire-9
Exploring a Need for a Cardiometabolic Disease Staging System as a Computerized Clinical Decision Support Tool: Qualitative Study

Aizhan Karabukayeva¹*, MPH; Jami L Anderson¹*, MSc, MBS; Allyson G Hall¹*, PhD; Sue S Feldman¹,2*, MEd, PhD, RN; Tapan Mehta³*, MSEE, PhD
¹Department of Health Services Administration, School of Health Professions, University of Alabama at Birmingham, Birmingham, AL, United States
²Department of Graduate Medical Education, Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States
³Department of Family and Community Medicine, Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States
*all authors contributed equally

Corresponding Author:
Aizhan Karabukayeva, MPH
Department of Health Services Administration
School of Health Professions
University of Alabama at Birmingham
1716 9th Avenue South
Birmingham, AL, 35294
United States
Phone: 1 2055155146
Email: ak1984@uab.edu

Abstract

Background: Although cardiometabolic diseases are leading causes of morbidity and mortality in the United States, computerized tools for risk assessment of cardiometabolic disease are rarely integral components of primary care practice. Embedding cardiometabolic disease staging systems (CMDS) into computerized clinical decision support systems (CDSS) may assist with identifying and treating patients at greatest risk for developing cardiometabolic disease.

Objective: This study aimed to explore the current approach to medical management of obesity and the need for CMDS designed to aid medical management of people living with obesity, at risk of being obese, or diabetic at the point of care.

Methods: Using a general inductive approach, this qualitative research study was guided by an interpretive epistemology. The method included semistructured, in-depth interviews with primary care providers (PCPs) from university-based community health clinics. The literature informed the interview protocol and included questions on PCPs’ experiences and the need for a tool to improve their ability to manage and prevent complications from overweight and obesity.

Results: PCPs (N=10) described their current approaches and emphasized behavioral treatments consisting of combined diet, physical activity, and behavior therapy as the first line of treatment for people who were overweight or obese. Results suggest that beneficial features of CDSS include (1) clinically relevant and customizable support, (2) provision of a comprehensive medical summary with trends, (3) availability of patient education materials and community resources, and (4) simplicity and ease of navigation.

Conclusions: Implementation of a CMDS via a CDSS could enable PCPs to conduct comprehensive cardiometabolic disease risk assessments, supporting clinical management of overweight, obesity, and diabetes. Results from this study provide unique insights to developers and researchers by identifying areas for design optimization, improved end user experience, and successful adoption of the CDSS.

(JMIR Form Res 2022;6(7):e37456) doi:10.2196/37456

KEYWORDS
cardiometabolic disease staging system; risk assessment; cardiometabolic disease; clinical decision support system; primary care; obesity; overweight; medical management
Introduction

Cardiometabolic diseases are leading causes of morbidity and mortality in the United States, including a wide array of diseases, typically beginning with insulin resistance and progressing later into a cluster of conditions that increase the risk of type 2 diabetes, stroke, and cardiovascular disease [1,2]. Being overweight (BMI ≥25 kg/m²) is associated with double the risk of developing cardiometabolic multimorbidity, while having mild and severe obesity (BMI ≥30 kg/m²) increases the risk 4 and 10 times, respectively [3]. However, current diagnostic categories that are based on standard BMI ranges defining overweight and obesity have high specificity but low sensitivity for identifying insulin resistance and cardiometabolic disease [4]. For example, with the current diagnostic categories, some individuals with overweight and obesity might not have cardiometabolic risk factors and may exhibit low rates of future diabetes and cardiovascular-related mortality; alternatively, some individuals who do not meet criteria for either metabolic syndrome or prediabetes exhibit risk of future diabetes [4]. Thus, risk assessments for cardiometabolic disease with greater sensitivity should be an integral component of medical practice, with tools to evaluate preventive and therapeutic options in patients at greatest risk for developing disease. Currently, there is no stratification of the population by level of obesity-related disease and mortality risk [5].

An estimated 42.5% of US adults aged 20 years and older are living with obesity, including 9.0% with severe obesity, and another 31.1% are overweight [6]. Because this group is at high risk of developing diabetes and other obesity-related complications, there is a need for risk stratification approaches to identify early those at highest risk and identify weight loss programs with appropriate treatment intensity. To provide appropriate medical management of obesity and facilitate the diabetes risk assessment of people with excess adiposity, a comprehensive staging system that establishes 5 stages of cardiometabolic disease risk—the cardiometabolic disease staging system (CMDS)—was developed [7,8]. This validated staging system is based on Adult Treatment Panel III metabolic syndrome risk factors and includes waist circumference, systolic and diastolic blood pressures, fasting and 2-hour blood glucose levels, triglycerides, and high-density lipoprotein cholesterol (HDL-C; Table 1) [4]. The purpose of this system is to help clinicians select treatment modality and intensity in the management of cardiometabolic diseases while balancing benefit and risk. Evidence demonstrates the CMDS has higher predictive and discriminative ability compared with other systems and relies on data typically collected during primary care visits; thus, it is more feasible to integrate into busy workflows of primary care providers (PCPs) [5].

<table>
<thead>
<tr>
<th>Stage</th>
<th>Descriptor</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 0</td>
<td>Metabolically healthy</td>
<td>No risk factors</td>
</tr>
<tr>
<td>Stage 1</td>
<td>One or two risk factors</td>
<td>Have 1 or 2 of the following risk factors: 1. High waist circumference (≥112 cm in men and ≥88 cm in women) 2. Elevated blood pressure (systolic ≥130 mm Hg and/or diastolic ≥85 mm Hg) or on antihypertensive medication 3. Reduced serum HDL-C (&lt;1.0 mmol/L or 40 mg/dL in men; &lt;1.3 mmol/L or 50 mg/dL in women) or on medication 4. Elevated fasting serum triglycerides (≥1.7 mmol/L or 150 mg/dL) or on medication</td>
</tr>
<tr>
<td>Stage 2</td>
<td>Metabolic syndrome or prediabetes</td>
<td>Have only 1 of the following 3 conditions in isolation: 1. Metabolic syndrome based on 3 or more of 4 risk factors: high waist circumference, elevated blood pressure, reduced HDL-C, and elevated triglycerides 2. Impaired fasting glucose (IFG; fasting glucose ≥5.6 mmol/L or 100 mg/dL) 3. Impaired glucose tolerance (IGT; 2-h glucose ≥7.8 mmol/L or 140 mg/dL)</td>
</tr>
<tr>
<td>Stage 3</td>
<td>Metabolic syndrome + prediabetes</td>
<td>Have any 2 of the following 3 conditions: 1. Metabolic syndrome 2. IFG 3. IGT</td>
</tr>
<tr>
<td>Stage 4</td>
<td>T2DM and/or CVD</td>
<td>Have T2DM and/or CVD: 1. T2DM (fasting glucose ≥126 mg/dL or 2-h glucose ≥200 mg/dL or on antidiabetic therapy) 2. Active CVD (angina pectoris or status post a CVD event such as acute coronary artery syndrome, stent placement, coronary artery bypass, thrombotic stroke, nontraumatic amputation due to peripheral vascular disease)</td>
</tr>
</tbody>
</table>

aHDL-C: high-density lipoprotein cholesterol.
bT2DM: type 2 diabetes mellitus.
cCVD: cardiovascular disease.
Vigilance in the management of modifiable risk factors is critical, given that people with overweight and obesity are at increased cardiovascular risk. Primary care settings, as familiar and accessible clinical venues for patients, are well positioned to screen people with overweight and obesity and recommend appropriate weight loss treatment plans to prevent complications and weight progression. Many studies found that the largest weight losses were achieved with high-intensity counseling by PCPs and referral of interested individuals to appropriate interventions [9-13]. However, a study of a nationally representative sample of adults aged 35 years and older found that, despite more adults reported being screened for obesity (78.6%) and of those screened, nearly 40% had a BMI of 30 kg/m² or higher (39.2%), only slightly more than one-half (53.5%) of obese adults screened reported receiving counseling about weight management [14]. Furthermore, BMI is the most preferred screening tool, though literature indicates it could be a poor indicator of cardiovascular disease and overall mortality risk [7,15]. Research finds BMI is not a good index of visceral fat, which is the basis of metabolic disorders associated with increased cardiovascular risk, whereas waist circumference might be superior as a risk assessment tool [16]. PCP-indicated practice improvements, helpful in treating and managing overweight and obesity, include better tools for early identification of risk and preventive treatment for those with multiple risk factors [11].

Providing CMDSS to PCPs via computerized clinical decision support systems (CDSS) may assist in stratifying the population by obesity-related disease risk and targeting those patients who are at greater risk for obesity-related complications. To the authors’ knowledge, this would be the first electronic health record–integrated CDSS that would incorporate CMDSS. Despite literature indicating CDSS may have a positive impact on provider performance and patient outcomes [17], evidence also indicates that CDSS rarely reach their full potential [18]. As with any innovation, user acceptance and integration within the clinical workflow are critical for successful uptake and routine use [19].

System analysis and design involve the process of planning, analyzing, designing, developing, implementing, and maintaining systems. A user-centered approach focusing on the user experience necessitates coordinated relationships between the system specialists, designers, and developers and the non-specialists and users with outcomes knowledge. The system development life cycle, when combined with the user experience life cycle, allows for that coordination to occur and has been shown to lead to better system adoption [20]. Figure 1 illustrates our conceptual model for system analysis and design of the CMDSS. This paper reports on the first 3 phases of each cycle: (1) plan and define, (2) analyze and research, and (3) design. As such, with the aim of involving users at key milestone stages of system development, this study explored the current approach to management of overweight and obesity and a need for the CMDSS system at the point of care to facilitate specificity in treatment modalities.

Figure 1. Conceptual model for system analysis and design.

Methods

To ensure we adhered to qualitative reporting standards, we followed the 32-time consolidated criteria for reporting qualitative studies (COREQ) checklist (Multimedia Appendix 1).

Sampling

Participant recruitment used convenience sampling where the research team coordinated with the medical director for primary care of a large academic medical center in the southeastern United States. Recruitment emails to potential candidates indicated the study purpose and invited participation. The number of participants was determined to be sufficient when saturation was reached (N=10) [21]. All participants were...
Data Collection

From August 2020 to January 2021, 10 semi-structured interviews were conducted by 2 research team members, consisting of the principal investigator (TM; male) and a graduate research assistant (AK; female). Two senior female researchers (AH and SF) with training in qualitative interviewing provided guidance and supervision. The interviewers did not have prior relationships with the participants. Only the interviewers and participants were present during data collection. The duration of the interviews varied between 30 minutes and 45 minutes and were conducted via a collaborative, cloud-based videoconferencing service at a mutually agreed-upon time. Interviews were transcribed verbatim by a commercial transcription company. The interview guide was informed by the literature review and included questions designed to (1) understand how PCPs manage overweight and obesity and facilitate prevention and management of diabetes and cardiovascular disease risk during a standard primary care visit and (2) explore PCP needs for CMDS and preferences for a CDSS (Multimedia Appendix 2). Broad, open-ended questions along with permissive prompts were used to facilitate each semi-structured interview. Prior to conducting interviews, the semi-structured interview guide was pilot tested with several providers to ensure questions were clear, generated in-depth discussion, were acceptable to participants, and resulted in usable information. Feedback from pilot testing was used to modify the wording, content, and order of the interview questions.

Ethical Considerations

All investigations were conducted in conformity with ethical principles of research. Consent for participation and interview recording was obtained verbally before each interview. This study was determined to be exempt by the University of Alabama at Birmingham Institutional Review Board (IRB Protocol Number 300003559).

Data Analysis

Transcribed interviews were coded using an inductive thematic analysis approach with NVivo 12 Plus (QSR International, Melbourne, Australia). To increase reliability and reduce bias, all transcripts were coded by 2 team members (AK and JA) with expertise in thematic analysis [22,23]. The analysis consisted of 2 phases: codebook development and codebook refinement. First, during open coding, coders examined an initial set of transcripts for categories (processes or events that share an attribute) of information related to our research questions. The second phase of our analysis focused on comparing and applying our initial codes to both existing and new data generated from subsequent interviews. This constant comparative analysis [24] across data sets allowed merging and clarifying codes. Following the initial coding process, research team members (AK and JA) discussed questions and discrepancies until 95% agreement was reached. Then, coders identified key points and recurring categories and themes that were central to the experience described by the participants. The process consisted of both coders dividing the text into semantic segments, labelling the segments with codes, together examining the codes for overlap and redundancy, and aggregating these codes into broader categories and themes [25].

Results

Sample Characteristics and Suggestions

We recruited 10 PCPs (7 physicians and 3 certified registered nurse practitioners) with practice experience ranging from 3 years to 43 years, with a mean of 12.2 years. Out of 10 respondents, 4 were male, and 6 were female. The most common practice-based barriers included lack of time and knowledge of resources, including access to evidence-based medical models and affordable community options. Considering the results of this study, 4 factors emerged as important for consideration in the development of a CDSS for metabolic conditions: (1) clinically relevant and customizable information delivery, (2) provision of comprehensive medical summary with trends, (3) availability of patient education materials and community resources, and (4) simplicity and ease of navigation. Table 2 describes the key suggestions voiced by the PCPs for future design of the CDSS to be successfully adopted.
Focus Not on Prevention But on Comorbidities

Almost all respondents reported that a significant portion of their patient populations was overweight, and they also noted that about 60% to 70% of patients had hypertension, diabetes, or other comorbidities. Even young populations presenting to primary care tended to have elevated BMIs or abnormal glucose levels. However, the respondents noted that they gave priority to management of the comorbidities rather than focusing on prevention and management of obesity. Respondents also noted they did not routinely use pharmacologic treatments for overweight or obesity but more to treat comorbidities, such as hypertension or elevated blood glucose levels.

BMI as a Main Diagnostic Measure

According to the respondents, BMI remains the primary tool for assessing obesity, as it is easy to access, is affordable to measure, and can conveniently be used to monitor weight changes. Additionally, participants responded that waist circumference measurement has not been integrated into routine practice. Patient risk factors associated with being identified or diagnosed as overweight or obese by their physician included higher BMI, family history, lifestyle, and habits. Respondents noted that they provided metabolic screening depending on patient’s BMI, including blood glucose and blood lipids levels.

Reliance on Lifestyle Modifications

Most of the respondents’ approaches to weight management were limited to assessing physical activity and assessing readiness for change, dietary habits, and expectations. The most common recommendations were to increase physical activity and dietary changes. Interestingly, half (5/10, 50%) of the

Table 2. Suggestions from primary care providers regarding preferrable clinical decision support system features.

<table>
<thead>
<tr>
<th>Suggestions</th>
<th>Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed of the information technology</td>
<td>“The other thing would be – does it run efficient? There are parts of Cerner that literally if you click the button, you’ re going be sitting there for 2 minutes just waiting, waiting, and waiting.” [Primary care physician, male]</td>
</tr>
<tr>
<td>Synthesis of available information</td>
<td>“I think what would be good is if you had a piece of software that could extract that [lab] data out of the record. And then you could click on a button at the top of the record, and it said ‘weight management’. If you click, it would have drop down algorithm and it was connected to the orders.” [Primary care physician, male]</td>
</tr>
<tr>
<td>Fit in the workflow</td>
<td>“So, whatever you come up with has to be something that’s integrated and uses the data that’s there, and gives you immediate feedback. It can’t be something that takes three minutes to enter the data.” [Primary care physician, male]</td>
</tr>
<tr>
<td>User-friendly with minimalist design</td>
<td>“So, ideally something self-contained, within the same page gives me kind of risk information and recommendations based of that, especially if it could be set up such that off of that page, I could directly order things. That would be amazing.” [Primary care physician, male]</td>
</tr>
<tr>
<td>Flexibility</td>
<td>“I think you definitely need to maintain the ability to customize or edit because, again, these are just sort of recommendations and sort of a part of the picture that the risk calculator gives you, but, you know, as long as you know, you could sort of edit to customize and individualize to a patient.” [Primary care physician, female]</td>
</tr>
<tr>
<td>Justification of treatment based on guidelines</td>
<td>“If there was something to standardize [management of] obesity and would give you a quantifiable number that puts them at a higher risk factor. So, if there was something that took in more either genetic versus biological markers that could be influential, I think that would be very useful and something that we would definitely want to implement and make it more of a standardization and not just an extra research tool.” [Primary care physician, male]</td>
</tr>
</tbody>
</table>

Obesity is always important, but it’s probably like number 10 on the list of concern. [Certified registered nurse practitioner, male]

We don’t measure it [waist circumference] in our clinic. We do have the BMI. So, the first point is BMI- this is all I look at because that’s what I have available, and it’s just a measure of numbers and the calculation. So, it’s easy. [Internist, female]

I look at their medical and family history. Like if they have diabetes in the family, then obviously that puts them at a higher risk automatically. Or if they have family members with hypertension. So, family history is very important for my understanding. [Internist, female]
respondents noted they did not have any formalized treatment plan to manage overweight or obesity and did not follow specific treatment guidelines. In addition, there was limited use of external sources of weight management support, with only few patients being referred to weight loss clinics, mainly due to limited coverage of services by health insurance companies. External resources frequently included a nutritionist and a commercial weight loss program (eg, Weight Watchers).

I do not have sort of very specific treatment guidelines. I’m not saying I’d be opposed to that. I just have my own practice at this time. It really will depend, because I’m trying to gauge a person’s willingness to change, so I will certainly ask some typical, open-ended questions about what have they tried in the past. It really becomes an individualized approach.  [Internist, male]

I would say exclusively exercise and diet. More recently, I [started] referring patients with BMI 30 or higher with multiple comorbidities to the weight loss management program.  [Primary care physician, female]

Lack of Knowledge About Referral Options in the Community

Respondents agreed that resources for intense lifestyle intervention and social support were important for the patients; however, respondents also noted the lack of knowledge about referral options in the community, including commercial-based programs. Because of the range in the socioeconomic status of their patient population, respondents expressed wanting point-of-care information about various affordable and convenient options that would be readily accessible and affordable for patient engagement.

I think there are a lot of resources out there but to be honest with you, I don’t think that we really know where the resources are. Everyone, probably, has their own little list of resources that they use, I think.  [Primary care provider, male]

I don’t think I necessarily have a good handle on it [the local resources]. I have certain things that I would say as a ‘go to’ that are probably out of date and missing a lot of some of the newer [resources].  [Certified registered nurse practitioner, female]

Lack of Patient Education Literature

One of the challenges voiced by respondents was lack of appropriate, “meet them where they are” weight management educational materials accessible for use at the point of care or after consultation.

Time during a visit is at a premium. In theory, our visits are 20 minutes, by the time the patient gets here, checks in, and triaged, I generally have about seven minutes out of 20 minutes to see a patient. Maybe a little bit more, sometimes a little less.  [Primary care physician, male]

The printed materials are not very good that we have available. They are not very helpful. That is why I don’t give them out very often.  [Primary care physician, female]

Need for CMDS

Need for a Risk Stratification Tool Embedded Into the CDSS

Almost all respondents (9/10, 90%) expressed an interest in having a CDSS that would incorporate diabetes and cardiovascular disease risk assessment and, based on the risks, outline a treatment plan. As respondents noted, the advantage of using a CDSS would be providing “legitimate justification” for a treatment plan with an assumption that patients understand their risk and the reason for the proposed treatment. To the authors’ knowledge, there is no decision support system available to assist providers in evidence-based weight loss treatment intensification. There is, however, a diabetes management protocol that has been developed but is not part of the electronic health records at this institution.

Currently, I have to pick up my phone, get on my coronary app and then put all the information. So, you could see where a tool like this that is incorporating the coronary risk score would be quite helpful built within Cerner. If it could even populate the data that we have with more recent blood pressure, that would be even more useful.  [Primary care physician, male]

Ideally, it would be something that I could just turn the computer monitor and show the patient, saying “Okay, well, this is why I’m recommending it. Your A1C is 5.9, up from 5.6 last year. Your cholesterol is up, your weight is up. So, this gives you a 17% chance of diabetes in the next two years. And these are the steps that we recommend...  [Primary care physician, male]

Diagnostic-Supported CDSS

Respondents noted a need for an CDSS that would consider diagnostics, such as relevant patient data and lab results. In addition, respondents indicated it would be useful to have access to clinically meaningful trends and track risk scores for complications. A majority of respondents manually calculate various risk scores, such as the 10-year Atherosclerotic Cardiovascular Disease (ASCVD) risk; therefore, embedding such a calculator in the CDSS could increase efficiency and reduce error.

If we had a good algorithm based on BMI and any potential risk factors that was easy to follow and implement, with good handouts and appropriate referral or community resources there, and if we could collate that information, I think it could be helpful.  [Primary care physician, male]

The main risk calculator that I use is the 10-year cardiovascular risk when I’m trying to decide if somebody should be on a statin. I just have on my phone and I just pull it up when I get their [patients’] labs back and plug in the numbers. Risk calculator
Incorporating Evidence-Based Practice

Respondents thought that having a CDSS that incorporates evidence-based clinical guidelines for management of obesity or overweight, both medical and behavioral, and that provides intervention recommendations would standardize and streamline the care provided and interventions suggested to their patients. The general idea was that such a system might help assist with managing patients with required tests, follow-up appointments, and preventive care.

I wouldn't say that I could speak for the whole clinic, I may just not be getting something that everyone else is doing. But we would be very open to having a tool that brings a standardization and also makes sure you're not overlooking anything and following a best practice guideline on initial management, and then also routine follow up. [Internist, female]

I think what would be most beneficial is having suggested treatment plans. It would help me to know that I am on the right track if I had a treatment plan that was suggested based upon their [patient] other chronic co-morbidities and their current A1C results or current blood sugar trends. [Certified registered nurse practitioner, female]

Ability to Have Resources to Make Referrals and Educate Patients

Respondents expressed a strong interest in information about accessible and affordable resources in the local community for patients struggling with overweight or obesity. They believed that it would enhance patient engagement and motivate patients toward behavior change. In addition, several respondents suggested that, if the CDSS had the means to efficiently provide appropriate educational materials to patients, it could improve the patient’s participation in their own care:

...having good patient-friendly handout material that was easy to attach in the patient portal in terms of a new diet, recommendations based on ADA or other more popular diets like the DASH (Mediterranean Diet) or other types of things based on the patient’s history. I do think that would be helpful. [Primary care physician, female]

Maybe [CDSS] gives you an option that you can click on, like option A “Would this person be interested in nutritional counseling?” , option B “Do you want to print this list of printouts to give them and present to them during your clinic visit?”, or “Would they want a referral to weight loss clinic?” I mean, it would be awesome if we had some way to refer people to some community resource near them where they could be contacted and offered some kind of like exercise class or a way to get into a walk group or something. [Certified registered nurse practitioner, female]

Discussion

Principal Findings

In this study, we sought to understand and capture user requirements for a system that evaluates the stage and severity of cardiometabolic disease that would be incorporated into a CDSS. The idea to involve intended users early in the design process is well supported in the literature [26] and results in aligning user expectations with the resulting functionality [27]. Currently, the PCPs’ approach to management of overweight and obesity largely focuses on treating comorbidities and counseling lifestyle modifications such as diet and exercise. There is limited use of medications to combat overweight and obesity. Although there are existing guidelines for obesity treatment [28] and related cardiometabolic conditions such as diabetes [29] and hypertension [30], our respondents were not consistently using them for diagnosis or treatment purposes.

Comparison With Prior Work

Findings from this study are consistent with a growing body of literature on how PCPs manage overweight and obesity, as well as on what CDSS features increase likelihood of its uptake. Turner et al [31] found, among a nationally representative sample of active health care providers, that (1) knowledge of physical activity and dietary guidelines was limited and (2) understanding of the appropriate initiation, intensity, and duration of pharmacotherapy was often inconsistent with evidence-based guidelines. Another study found that PCPs were least likely to say they would prescribe medication or refer a patient to counseling [32]. Regarding CDSS features, a systematic review by Groenhof et al [33] found that design and usability were important drivers behind the success, noting that information should be displayed all at once and at one glance. In addition, the lack of insight into the automated computation and source of information decreased user satisfaction. Further, the most recent systematic review by Kouri et al [34] identified important CDSS features that significantly predict uptake, such as avert the need for provider data entry by mining patient data from within electronic health record systems to inform CDSS.

Implication for Practice

An important consideration would be providing easy access to the latest evidence-based clinical standards and protocols by embedding them in the CDSS. As a first step, CDSS could include measurement tools to perform a diagnostic evaluation based on evidence-based guidelines. If currently only BMI is considered, a more comprehensive evaluation must include additional measures such as measurement of waist circumference. Further, for the treatment, the physicians should be able to obtain clinical decision support by using CDSS to analyze pertinent information about the patient’s current clinical condition, including information about medication, lab results, and treatment compliance. Given support for the CDSS among our sample, we propose a design of a CDSS that provides suggestions for treating the primary and augmenting medications with explanations. For the purposes of follow-up, the CDSS should have reminders to ensure the important considerations are not overlooked. Moreover, it could also recommend and...
display when the patient should return for a visit. All entries should be automatically stored, providing electronic documentation and record keeping, thus providing access to complete patient information. Overall, information about a patient’s demographic characteristics and other clinical records should be accessible by a single click.

Implications for Development

Results from this study were used to better understand user requirements within a parallel system analysis and design framework (see Figure 1), the importance of which was to ensure the voice of the user was adequately and accurately represented [20]. In this phase, we present the conceptual framework with the findings applied as high-level categories (see Figure 2). These categories and the details behind them as presented throughout this study will be used to inform the evaluation.

Figure 2. The conceptual framework with the findings as high-level categories.

Limitations

There are several limitations of this study. First, the 10 participants, predominantly non-Hispanic White (9/10, 90%), represented PCPs at a large academic center in the southeast. A small homogenous sample size could potentially limit the generalizability of our findings, and we recognize the need to add breadth and depth to this participant sample as development ensues. In addition, the strength of the study was our consideration of assuring intercoder reliability. Thus, we feel confident that we are correctly representing the voices of our participants.

Conclusion

Implementation of a CMDS system in the form of a CDSS could be used as a risk assessment tool that also provides risk-based and evidence-based treatment or program recommendations to better manage overweight and obesity and prevent diabetes. Results from this study provide unique insight to developers and researchers to identify areas for design optimization for improved end user experience to ensure successful adoption of the CDSS.

Acknowledgments

The study described was supported by a pilot and feasibility award to TM from the University of Alabama at Birmingham (UAB) Diabetes Research and funded by P30DK079626 from the National Institute of Diabetes and Digestive and Kidney Diseases. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Diabetes and Digestive and Kidney Diseases or the National Institutes of Health.

The authors wish to thank UAB Prime Care leadership and providers who participated and facilitated this study.

Conflicts of Interest

None declared.
References


Abbreviations
- ASCVD: Atherosclerotic Cardiovascular Disease
- CDSS: computerized clinical decision support system
- CMDs: cardiometabolic disease staging
- COREQ: consolidated criteria for reporting qualitative studies
- HDL-C: high-density lipoprotein cholesterol
- PCP: primary care provider
- UAB: University of Alabama at Birmingham

https://formative.jmir.org/2022/7/e37456
JMIR Form Res 2022 | vol. 6 | iss. 7 | e37456 | p.21
(page number not for citation purposes)
Exploring a Need for a Cardiometabolic Disease Staging System as a Computerized Clinical Decision Support Tool: Qualitative Study

Karabukayeva A, Anderson JL, Hall AG, Feldman SS, Mehta T

JMIR Form Res 2022;6(7):e37456
URL: https://formative.jmir.org/2022/7/e37456
doi: 10.2196/37456
PMID: 35776499

© Aizhan Karabukayeva, Jami L Anderson, Allyson G Hall, Sue S Feldman, Tapan Mehta. Originally published in JMIR Formative Research (https://formative.jmir.org), 01.07.2022. 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

Improvements in Depression Outcomes Following a Digital Cognitive Behavioral Therapy Intervention in a Polychronic Population: Retrospective Study

Aarathi Venkatesan¹, PhD; Benjamin Forster¹, BA; Prasanna Rao¹, BS, MS; Melissa Miller¹, BA, MS, MBA; Michael Scahill¹, MD, MBA
Vida Health, San Francisco, CA, United States

Corresponding Author:
Aarathi Venkatesan, PhD
Vida Health
100 Montgomery St, Ste 750
San Francisco, CA, 94104
United States
Phone: 1 14159891017
Email: aarathi.venkatesan@vida.com

Abstract

Background: Digital mental health interventions have shown promise in reducing barriers to effective care for depression. Depression and related mental disorders are known to be highly comorbid with common chronic physical conditions, such as obesity and type 2 diabetes. While some research has explored the interaction dynamics of treating populations living with both mental and physical disorders, very little is known about such dynamics in digital care.

Objective: We aimed to examine the effectiveness of a 12-week, therapist-supported, app-based cognitive behavioral therapy program in improving symptoms of depression and anxiety. The studied population included adults with a heavy burden of chronic physical disease, including obesity and type 2 diabetes.

Methods: A total of 1512 participants with at least moderate depression were enrolled. The treatment cohort consisted of 831 (54.96%) participants who completed a follow-up assessment. The program included structured lessons and tools (ie, exercises and practices) and offered one-on-one weekly video counseling sessions with a licensed therapist for 12 weeks and monthly sessions thereafter. The clinically validated 8-item Patient Health Questionnaire (PHQ-8) and the 7-item Generalized Anxiety Disorder scale (GAD-7) were used to assess depression and anxiety, respectively. Linear mixed-effects modeling was employed to examine changes in depression and anxiety over time. Given correlation among various measures of program usage, a composite variable for depth of usage was used to analyze the correlation between usage and changes in depressive symptoms. Body weight changes from baseline were assessed primarily with digitally connected scales.

Results: Out of 831 participants in the treatment cohort, 74.5% (n=619) showed a clinically significant reduction in depressive symptom severity after 12 weeks, where follow-up PHQ-8 scores had shifted downward by at least one diagnostic category. In total, 67.5% (n=561) of the participants showed a reliable improvement in PHQ-8 scores as measured by the reliable change index. There was an average reduction of 5.9 (SD 5.2) points (P<.001) between baseline and follow-up. Greater program usage was correlated with greater likelihood of reliable improvement in depressive symptoms (odds ratio 1.3, 95% CI 1.1-1.5; P=.002). An exploratory analysis of body weight changes with a multilevel, mixed-effect model suggested that reliable improvement in depressive symptoms at follow-up was associated with significantly greater weight loss at 9 months (β=−1.11, P=.002).

Conclusions: The results provide further support that digital interventions can support clinically meaningful improvements in depression. Some form of synergy in treatment of comorbid depression and obesity or diabetes could be studied in future research. The study was limited by postintervention participant attrition as well as the retrospective observational study design.

(JMIR Form Res 2022;6(7):e38005) doi:10.2196/38005

KEYWORDS
depression; anxiety; CBT; digital mental health intervention; cognitive behavioral therapy; digital health; obesity; diabetes; mental health
Introduction

In 2016, mental health disorders affected more than 1 billion people worldwide [1]. Among mental health disorders, clinical depression carries a lifetime risk of 15% to 18% [2] and has the strongest association with disability-adjusted life years, a global benchmark of disease burden [1,2]. Depression is also the psychiatric condition most strongly associated with suicide [3]. The relationships among clinical depression, well-being, and health care costs have been well documented, with costs of care for patients living with depression higher across spending categories [4,5].

It has been estimated from nationally representative surveys that the prevalence of depressive symptoms in the United States increased 3-fold during the COVID-19 pandemic, with the majority of those affected already having been at elevated risk [6]. Simultaneously, the pandemic has accelerated the adoption of digital technology to deliver health care, particularly for mental health, and highlighted the promise of remotely delivered health services when traditional models may not have been available or accessible [7].

Digital mental health interventions (DMHIs) are potentially scalable and effective treatment solutions for mental health. Although research evaluating the effectiveness of DMHIs notes significant heterogeneity in terms of study design, clinical model, and intervention characteristics, meta-analytic reviews of smartphone-based interventions have observed a clinically significant treatment effect compared to both wait-list and active controls for depression and anxiety symptoms [8-10].

There is considerable research consensus that cognitive behavioral therapy (CBT) is a highly effective therapeutic modality for the treatment of clinical depression and anxiety because of relatively short intervention durations with strong outcomes [11,12]. Specifically, CBT-based digital interventions were associated with a significantly greater reduction in depression compared to non-CBT therapeutic approaches (eg, mindfulness, psychoeducation, and mood monitoring) [9]. In the context of DMHIs, digital CBT interventions appear to be equivalent to face-to-face interventions in terms of treatment efficacy [13-16]. Furthermore, a recent assessment of digital CBT offerings that evaluated patient preference, cost savings, and clinical benefit concluded that a provider-guided digital CBT program was likely the most effective approach for depression and anxiety compared to self-guided and face-to-face interventions [17].

In addition to evaluating the efficacy of CBT-based digital interventions, there has been growing research interest in exploring their mechanisms, particularly in the role of program engagement in DMHIs. Recently, Chien et al [18] employed a machine learning–based approach to explore patterns of engagement with a digital CBT program and their relationship with improvements in depression and anxiety. While the study did identify five distinct classes of program engagement, from low to high engagement, all classes were all positively associated with clinical improvement. Higher engagement classes were associated with greater improvement in depressive symptoms, so a possible dose-response effect was suggested. However, it has been observed that user engagement can be difficult to define and measure [19]. Where usage generally measures observable actions, engagement implies some subjective experience of the digital intervention with a focus on the quality of the experience [20,21]. Furthermore, as Torous et al [22] note, usage over time, a commonly used measure of engagement, cannot distinguish a user whose mental health needs have been met by an app with a single instance of usage from one who needs repeated access for support.

One app feature that has been previously operationalized as a measure of DMHI engagement is lesson completion [23]. Lessons are typically brief pieces of evidence-based content intended to support a skill or health habit. They can also serve a crucial function in mental health interventions, digital or otherwise. CBT, in particular, emphasizes the practice of skills in between sessions by way of “homework assignments” or lessons in order to reinforce practices such as cognitive reappraisal [24]. Kazantzis et al [25] observed a stronger treatment effect size for CBT-based therapeutic interventions that incorporated homework when compared to interventions without a homework component. Lesson completion has also been used as a measure of treatment adherence and has shown a positive association with improvements in depression and anxiety [17,25-29]. It has been noted that the ubiquity of smartphones and the flexibility of the technology enables access to extensive content, on demand or needs based, that is difficult to replicate in traditional delivery models [28,30]. Lastly, an often-overlooked dimension of clinical depression, particularly in the context of DMHIs, is its frequent co-occurrence with chronic health conditions, such as type 2 diabetes and obesity [31-33]. Among other aspects, depression is known to hamper self-care and medication adherence [33]. Cross-sectional research suggests a 1.18 times greater likelihood of depressive symptoms in individuals with obesity than in those without [34]. Similarly, the risk of prediabetes and related measures appears to be markedly elevated among those newly diagnosed with depression [35,36]. This relationship tends to be stronger among women [34,35]. In their meta-analytic review of collaborative care to treat depression and diabetes in tandem, Atlantis et al [31] observed significant improvements in both depression and glycemic management with tandem treatment. Another systematic review showed that psychological interventions tailored for people with diabetes were effective in improving both glycemic management (ie, hemoglobin A1c [HbA1c]) and elevated diabetes distress [37]. There are indications that the relationship between chronic conditions and depression are bidirectional, with an increased incidence of diabetes among those with diagnosed depression [38-40]. Many app-based health interventions have a singular health focus (eg, depression, weight, or diabetes management), making it more challenging to fully understand the impact and the influence of co-occurring conditions on treatment adherence and health outcomes [41].

In sum, research suggests that CBT-based DMHIs are feasible and effective solutions for depression and anxiety disorders [42-44]. There are some preliminary indications that the extent of program usage may influence treatment adherence and
outcomes [45]. There remain open research questions on how intervention type, mode of delivery, and mechanisms may affect treatment outcomes. There has also been less focus on DMHIs in populations living with co-occurring depression and chronic health conditions.

In this study, we evaluated the Vida CBT Program for moderate depression and anxiety in a polychronic adult population, that is, one with a high prevalence of chronic health conditions. Vida Health is a Health Insurance Portability and Accountability Act–compliant, app-based platform for management of both mental and physical health that combines tailored content with counseling by licensed therapists and other health education specialists. The platform is available directly to consumers or as a benefit from select employers and health plans. The primary objective of this study was to assess changes in depression following a 12-week digital CBT program: the Vida CBT Program. We hypothesized that participants who completed the program would show a reduction in depressive symptoms and that measures of program usage would be positively associated with these improvements. As part of a preliminary exploratory analysis, we also evaluated changes in weight among participants concurrently enrolled in a Vida physical health program. We suspected that improvements in depression would be positively associated with stronger weight loss.

**Methods**

**Study Design**
This study used a single-arm, retrospective design to evaluate changes in depression and anxiety following the Vida CBT Program.

**Ethics Approval**
The study protocol and informed consent statement was reviewed and approved by the Western Institutional Review Board Inc (protocol No. 20192591), an independent institutional review board. All data were fully anonymized prior to data analysis. Informed consent statements were sent to participants upon enrollment in the Vida CBT Program.

**Measures**
The 8-item Patient Health Questionnaire (PHQ-8) was used to assess severity of depressive symptoms. The following standard scoring cutoffs were applied to classify depressive symptom severity: 0 to 4 (asymptomatic or minimal), 5 to 9 (mild), 10 to 14 (moderate), 15 to 19 (moderately severe), and 20 or higher (severe). Anxiety symptoms were assessed using the 7-item Generalized Anxiety Disorder scale (GAD-7) using the following standard score cutoffs to classify anxiety symptom severity: 0 to 4 (asymptomatic or minimal), 5 to 10 (mild), 11 to 17 (moderate), and 18 or higher (severe). The PHQ-8 and GAD-7 are widely used in clinical settings and have shown robust reliability and validity [39,40]. The assessments were administered in the app automatically at program start, week 6, week 12, and every 3 months thereafter for up to 1 year. Participants were encouraged to complete the survey on the day of receipt but had the option to complete the assessment at any point during the 2 weeks following receipt, after which the survey would disappear until the next assessment time point. Therapists also had the option of sending the instrument to the participant at any point during the intervention as they deemed clinically appropriate.

Body weight was a secondary outcome measure in this study. Weight outcomes were either recorded via a connected wireless scale or self-reported by the participant using a logging tool available in the Vida Health app. Participants could either sync their personal wireless scale to the app or order a digitally connected scale via the app.

**Study Sample and Recruitment**
The study was open to adults, 18 years of age or older, who were fluent in English and had access to a smartphone or tablet. Participants were recruited between September 2019 and January 2021 using a combination of emails, mailers, and phone outreach efforts. The Vida Health app is available to individuals across the United States and can be downloaded from the Apple App Store or Google Play. Participants were drawn from several insurance plans and employers for whom Vida Health was offered as a covered benefit.

Upon downloading the app, participants completed a brief intake questionnaire that included name, contact information, basic demographics (ie, age, gender, height, and weight), and existing health conditions. Participants were offered a variety of health domains to focus on. Some participants chose to begin the Vida CBT Program directly. Others began in the Vida digital programs for weight loss or diabetes management described previously [46,47]. Participants received and completed an intake PHQ-8, and those with a score of 10 or greater (moderate depression) were included in the study. All participants also completed a baseline GAD-7 assessment. Eligible participants were paired with a licensed therapist based on their state of residence and preferred times for consultations. Therapists were mental health professionals working for Vida Health, licensed by their state’s respective licensure board. Participants with severe depression (PHQ-8 score ≥20) or severe anxiety (GAD-7 score ≥15) were ineligible for the Vida CBT Program and were referred to alternative sources of care. Additional exclusion criteria included eating disorders, substance use disorder, suicidality, homicidality, acute posttraumatic stress disorder, and episodes of mania or psychosis. Participants who presented with any of the above symptoms during the intervention were referred for care outside of Vida Health. Participants who chose to enroll initially in weight loss or diabetes management programs and were then screened into the Vida CBT Program had the option to continue both programs simultaneously.

**Therapeutic Approach and Intervention**
A fundamental focus of CBT is to address maladaptive thinking patterns by understanding the associations among thoughts, emotions, and behaviors [48]. As part of the Vida CBT Program, participants received structured multimedia (ie, audio, video, or text) lessons, activities, and practices within the app. Based on core CBT principles, such as guided discovery and conscious re-evaluation, these lessons were designed to increase awareness of one’s thinking patterns and support the practice of alternative, adaptive thoughts [24]. Ahead of their initial consultation with a Vida therapist, participants completed an informed consent
form for psychotherapy that detailed their rights to confidentiality and limits to confidentiality, including mandated reporting requirements as stipulated by the therapist’s licensing board and state regulations. At the initial consultation, therapists performed a comprehensive biopsychosocial assessment that included a review of previous treatments and diagnoses, current presenting problem, and symptoms. Participants could communicate with their therapist using live video or audio consultations as well as with asynchronous messaging in the app.

Following the intake, therapists documented their initial diagnostic impressions and developed individualized treatment plans with short-term and long-term goals. Participants were offered weekly video or audio consultations with their therapist for the first 12 weeks and shorter monthly follow-up sessions thereafter for up to 1 year. Each therapist consultation comprised setting goals and homework activities for the upcoming week along with a review of strategies for further cultivating concepts and skills learned from earlier sessions. In between sessions, participants could complete assigned homework and use a thought tracker in the app. Lessons, once shared, remained available in the app for the participant to review and revisit concepts. The thought tracker feature allowed participants to record current thoughts using a Likert scale to assist in identifying, evaluating, and restructuring distorted thought patterns. In program weeks 10 through 12, therapists worked with participants to create a Wellness Recovery Action Plan intended to support maintenance of acquired skills and improved functioning and to prevent relapse [49]. Select screenshots from the program are shown in Figure 1.

**Figure 1.** Screenshots from the Vida CBT Program. CBT: cognitive behavioral therapy.

Participants who were concurrently enrolled in a program for chronic disease management also worked with an additional Vida provider (ie, a certified health coach or registered dietitian) who could collaborate on care with the relevant therapist. As in the Vida CBT Program, they were offered synchronous consultations weekly for up to 12 weeks and monthly thereafter for up to 1 year. All providers received extensive training on motivational interviewing, an approach that leverages improving perceived self-efficacy and autonomy to facilitate healthy behavior change [50]. Additionally, participants received app content covering topics such as nutrition, exercise, and medication adherence. All content was informed by evidence-based research and literature on health behavior change as described previously [46,47,51].

### Statistical Plan

Change in depression, based on PHQ-8 scores, between baseline and follow-up was the primary dependent variable. We performed a paired, 2-tailed *t* test to assess if there was a significant change in PHQ-8 scores from baseline. We used a second paired *t* test to evaluate changes in anxiety scores among participants who scored in the moderate anxiety range at baseline (ie, GAD-7 score ≥11). Mean normalization was applied to all continuous predictors (eg, age). A Boolean variable was created for the presence of co-occurring anxiety (1 = baseline GAD-7 score ≥11). Gender was also coded as a binary variable (1 = female). Since we could not assume that all therapists were equally effective, all regression analyses were conducted using a cluster-robust approach with therapist as a cluster group variable [52]. The reliable change index (RCI) for depression and anxiety was also computed. RCI, a commonly used measure in psychometrics, is the ratio of the difference in pre-post assessment scores to the standard error of measurement [47,48]. An RCI score of 1.96 or higher (ie, 1 above the 95% CI) is regarded as an indication of reliable, statistically meaningful change [53].

To explore the interaction between program usage and changes in depression, we tabulated five program features that broadly encompass the program experience: number of therapist consultations, number of messages sent to the therapist, number of “core” lessons opened, number of thoughts logged, and total number of content pieces opened. Core lessons were those related to fundamental CBT concepts (eg, cognitive restructuring, behavioral activation, and techniques for addressing maladaptive thinking) [24]. Data exploration revealed a right skew for each of the usage factors, that is, a subpopulation of participants who used the app features quite extensively. In order to retain these heavy app users, but to limit their influence on downstream analyses, all usage factors were right-winsorized at the 99th percentile [54]. The treatment cohort showed a notable female predominance, a finding not unusual...
in studies of mental health service use in the United States [55-57]. In this study, in order to adjust for this, the established technique of oversampling was employed, wherein 715 participants were drawn at random from the 116 participants who did not identify as female [58,59].

Finding, as expected, that features of program usage were correlated, we constructed a composite variable to limit collinearity. Participants were assigned an ordinal variable from 0 to 5, where 1 unit was assigned for each usage feature in which the participant was above the 25th percentile. Thus, someone with a depth-of-usage score of 5 had activity above the 25th percentile in each of the usage factors, while a score of 0 indicated usage within the 25th percentile across all usage features. We chose the 25th percentile as it empirically separated users into high and low usage categories without selecting only for the right tail of usage as alluded above. With reliable change as the binary dependent variable (1 = reliable improvement in depression symptoms), a cluster-robust logistic regression evaluated the relationship depth of usage and improvement in depression symptoms. Controls included therapist cluster, baseline PHQ-8 score, gender, presence of anxiety, and age.

A supplementary analysis evaluating changes in weight outcomes among participants concurrently enrolled in the Vida weight loss program was conducted. Analysis was restricted to participants who had been enrolled for at least 6 months and had logged weight at least twice in that period. A linear, mixed-effects model was used to address potential heterogeneity in the frequency and number of weight logs by participants and provider-level differences [60]. Percent change in weight from baseline was regressed on the following fixed factors: program time (in months), baseline BMI category (obesity or overweight), and reliable change (1 = reliable improvement in PHQ-8 scores; 0 = no reliable improvement) [61]. Participants and providers were specified in the model as random factors.

All data preparation and analyses were performed using Python (version 3.7.9; Python Software Foundation) and Stata (version 16.1; StataCorp LLC).

## Results

### Overview

A total of 1512 participants enrolled in the Vida CBT Program between September 2019 and January 2021. A schematic of the participant flow is presented in Figure 2. Of those, 54.96% (n=831) had a follow-up PHQ-8 assessment between weeks 6 and 12 and were considered part of the treatment cohort. Analyses evaluating changes in depression scores from baseline were restricted to the treatment cohort. The remainder of the study cohort, equal to 45.04% (n=681) of the participants, failed to complete a follow-up PHQ-8 during the assessment window. These participants were excluded from the primary analyses; however, we performed a supplementary intention-to-treat (ITT) analysis to evaluate overall changes in PHQ-8 scores from baseline across the entire cohort. Among those without a valid follow-up, 71.5% (487/681) did not have any assessment after their baseline. For these members, a baseline carryforward approach was employed. The remaining 28.5% (194/681) of the participants completed an assessment before finishing 6 program weeks. For these members, a last-value carryforward approach was applied.

Baseline characteristics of the study cohort are reported in Table 1. In addition to depression, the entire study cohort self-reported living with at least one chronic physical health condition (ie, type 2 diabetes, cardiovascular disease, or obesity). There were no significant differences in baseline PHQ-8 scores between the treatment and incomplete groups at baseline ($t_{1510}=1.2$, $P=.22$). There was a significantly lower rate of comorbid anxiety, defined as a GAD-7 score of 11 or higher, among participants in the treatment cohort ($\chi^2_1=19.5$, $P<.001$). A 2-tailed chi-square analysis indicated that there were significantly more women in the treatment cohort compared to the program noncompleter group ($\chi^2_1=7.1$, $P=.01$). There were also more participants concurrently enrolled in a health coaching program for a chronic condition in the treatment cohort ($\chi^2_1=171.1$, $P<.001$). Lastly, we observed a significant average difference in age of 2.3 years between the treatment cohort (mean 48.5, SD 11.4 years) and noncompleters (mean 46.2, SD 12.4 years; $t_{1510}=-3.8$, $P<.001$).
**Principal Results**

Out of 831 participants in the treatment cohort, 74.5% (n=619) showed a clinically significant reduction in depressive symptom severity in 12 weeks, where follow-up PHQ-8 scores had shifted downward by at least one diagnostic category. A total of 67.5% (n=561) of the treatment cohort participants showed a reliable improvement in PHQ-8 scores as measured by the RCI. There was an average reduction of 5.9 (SD 5.2) points between baseline and follow-up (Table 2). A 2-tailed, paired t test revealed a significant reduction in depression severity (β=−1.7, P<.001). That is, greater baseline depression severity was, unsurprisingly, associated with greater reduction of depression at follow-up. We observed that the co-occurrence of anxiety was associated with a smaller reduction in depression scores (β=1.33, P<.001). In supplementary analyses, a 2-tailed, paired t test including the entire ITT cohort, using carryforward as above, indicated a slightly attenuated but still significant reduction in PHQ-8 scores at follow-up (mean −3.6, SD 4.99; t_{1511}=29.1, P<.001).

Among the 478 participants with moderate anxiety or higher at baseline (ie, GAD-7 score ≥11), 89.7% (n=429) provided a follow-up assessment between weeks 6 and 12. A paired t test revealed a significant average reduction in anxiety scores of 6.1 (SD 5.4) points from baseline (β=−2.29, P<.001; Table 2). Additionally, 57.7% (n=276) of these participants had a reliable improvement in anxiety scores from baseline. A cluster-robust linear regression showed that higher baseline GAD-7 scores were associated with a smaller reduction in anxiety scores (β=−1.33, P<.001).

---

Table 1. Demographic characteristics of the treatment and intention-to-treat study cohorts.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Incomplete data (n=681)</th>
<th>Treatment (n=831)</th>
<th>Total (N=1512)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants (N=1512), n (%)</td>
<td>681 (45.0)</td>
<td>831 (55.0)</td>
<td>1512 (100)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>552 (81.1)</td>
<td>715 (86.0)</td>
<td>1267 (83.8)</td>
</tr>
<tr>
<td>Male</td>
<td>127 (18.7)</td>
<td>113 (13.6)</td>
<td>240 (15.9)</td>
</tr>
<tr>
<td>Not disclosed</td>
<td>2 (0.3)</td>
<td>3 (0.4)</td>
<td>5 (0.3)</td>
</tr>
<tr>
<td>Age in years, mean (SD)</td>
<td>46.2 (12.4)</td>
<td>48.5 (11.4)</td>
<td>47.5 (11.9)</td>
</tr>
<tr>
<td>Baseline PHQ-8a score, mean (SD)</td>
<td>14.6 (3.3)</td>
<td>14.4 (3.4)</td>
<td>14.5 (3.3)</td>
</tr>
<tr>
<td>Has anxiety (GAD-7b score ≥11), n (%)</td>
<td>467 (68.6)</td>
<td>478 (57.5)</td>
<td>945 (62.5)</td>
</tr>
<tr>
<td>Enrolled in physical health program, n (%)</td>
<td>360 (52.9)</td>
<td>697 (83.9)</td>
<td>1057 (69.9)</td>
</tr>
<tr>
<td>Chronic physical health condition, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obesity</td>
<td>554 (81.4)</td>
<td>723 (87.0)</td>
<td>1277 (84.5)</td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>367 (53.9)</td>
<td>503 (60.5)</td>
<td>870 (57.5)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>316 (46.4)</td>
<td>468 (56.3)</td>
<td>784 (51.9)</td>
</tr>
</tbody>
</table>

---

*PHQ-8: 8-item Patient Health Questionnaire; depressive symptom severity by score is classified as follows: 0 to 4 (asymptomatic or minimal), 5 to 9 (mild), 10 to 14 (moderate), 15 to 19 (moderately severe), and 20 or higher (severe).*

*GAD-7: 7-item Generalized Anxiety Disorder scale; a score of ≥11 indicates moderate to severe anxiety.*
were associated with greater reductions in anxiety scores at follow-up ($\beta=-3.5$, $P<.001$). Age, gender, and baseline PHQ-8 scores were not significantly associated with anxiety score changes.

Table 2. Estimated marginal means of PHQ-8 and GAD-7 scores at baseline and follow-up.

<table>
<thead>
<tr>
<th>Assessment type and time point</th>
<th>Score, estimated marginal mean (bootstrapped 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PHQ-8</strong></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>14.4 (14.2-14.6)</td>
</tr>
<tr>
<td>Follow-up (12 weeks)</td>
<td>8.5 (8.1-8.8)</td>
</tr>
<tr>
<td><strong>GAD-7</strong></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>14.8 (14.5-15.0)</td>
</tr>
<tr>
<td>Follow-up (12 weeks)</td>
<td>8.8 (8.3-9.3)</td>
</tr>
</tbody>
</table>

aPHQ-8: 8-item Patient Health Questionnaire; depressive symptom severity by score is classified as follows: 0 to 4 (asymptomatic or minimal), 5 to 9 (mild), 10 to 14 (moderate), 15 to 19 (moderately severe), and 20 or higher (severe).

bGAD-7: 7-item Generalized Anxiety Disorder scale; anxiety symptom severity by score is classified as follows: 0 to 4 (asymptomatic or minimal), 5 to 10 (mild), 11 to 17 (moderate), and 18 or higher (severe).

Program Usage Outcomes

Program usage was evaluated in the treatment cohort across a set of five program features: number of therapist consultations, number of messages sent to the therapist, number of times core lesson content was accessed, number of thoughts logged via the thought tracking tool, and number of program-related content cards viewed. As expected, there were notable correlations among these program usage factors, with Pearson coefficients ranging from 0.17 to 0.43 (Table 3).

In order to limit collinearity while evaluating the association between program usage and changes in depressive symptoms, we created a composite feature of overall depth of usage. For each usage feature, a score of 1 was assigned when activity for that specific feature was above the 25th percentile of the distribution. We then summed across each of the factors. Thus, depth-of-usage scores could range from 0 to 5, with a score of 5 indicating usage in excess of the 25th percentile of activity across the features, and a score of 0 indicating usage below the 25th percentile across the features. As noted above, the data were then resampled to adjust for the female predominance in the treatment cohort. From the resulting balanced cohort of 1430 participants, a cluster-robust logistic model controlling for age, gender, co-occurrence of anxiety (ie, baseline GAD-7 score $\geq 11$), and baseline PHQ-8 score revealed a significant association between depth of usage and likelihood of reliable improvement in depression scores at follow-up (odds ratio [OR] $1.3$, 95% CI 1.1-1.5; $P=.002$). In other words, greater usage across the platform was associated with improvement in depression symptom severity. The summary statistics for the features of program usage were as follows: mean number of therapist consultations was 7.04 (SD 3.4), mean number of messages sent to the therapist was 33.2 (SD 46.9), mean number of times core lesson content was accessed was 39.2 (SD 33), mean number of thoughts logged via the thought tracking tool was 13.6 (SD 16.5), and mean number of program-related content cards viewed was 8.35 (SD 12.4).
Table 3. Correlation analysis (Pearson r and 2-tailed P value) among the features of program usage (n=831).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Consults</th>
<th>Messages</th>
<th>Lessons opened</th>
<th>Thoughts logged</th>
<th>Content viewed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consults</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>1</td>
<td>0.29&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.26&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.26&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.17&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001&lt;sup&gt;b&lt;/sup&gt;</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Messages</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>0.29&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>0.33&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.43&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.31&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>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Lessons opened</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>0.26&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.33&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>0.44&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.35&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>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Thoughts logged</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>0.26&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.43&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.44&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>0.28&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>—</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Content viewed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>0.17&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.31&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.35&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.28&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
</tr>
<tr>
<td>P value</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>The correlation is significant at a significance level of .05 (2-tailed).

<sup>b</sup>Not applicable.

**Exploratory Weight**

We performed additional analyses to explore changes in body weight and diabetes management among participants simultaneously enrolled in the Vida CBT Program and a physical health program.

For body weight, in order to provide adequate time for meaningful clinical change, the analysis was restricted to the 595 participants with overweight or obesity (ie, BMI ≥25) who had been enrolled for at least 6 months and had logged weight at least twice in the period with the latest weight coming at least 2 months after the initial enrollment. Out of these participants, 86.6% (n=515) had a baseline BMI indicative of obesity and 13.4% (n=80) had a baseline BMI indicative of overweight. This population logged a total of 34,469 body weight values. Out of these entries, 83.17% (n=28,667) were logged using a wireless connected scale, and the remainder were entered manually by the participant. The majority of the participants (n=387, 65.0%) had logged their most recent weight in the ninth program month. The most recent weight was logged at least in the eighth program month for 78.8% (n=469) of the participants and at least in the sixth program month for 89.2% (n=531) of them.

To account for heterogeneity in the number and frequency of weight logs among participants, as well as to handle the nested structure of the data (ie, weight observations nested within participants who were nested within providers), a multilevel mixed-effects model was used to evaluate changes in body weight across participants. Fixed factors in the model were time (program month), baseline BMI category (obesity or overweight), and reliable change in depressive symptoms (1 = reliable improvement in PHQ-8 scores; 0 = no reliable improvement). Random factors included participants and providers. We observed a significant effect of program time on percent weight loss (β=−0.37, P<.001). Baseline weight had a small but significant inverse association with weight loss such that higher baseline weight was associated with greater percent weight loss (β=−0.01, P=.002). As shown in Figure 3, reliable improvement in depression symptoms at follow-up was associated with significantly greater weight loss at 9 months (β=−1.11, P=.002).
Discussion

Principal Findings

The aim of this retrospective study was to evaluate the therapist-supported, digitally delivered Vida CBT Program for the treatment of moderate depression in an adult population. The treatment cohort of 831 participants with baseline PHQ-8 scores of 10 or greater were enrolled in the digital intervention and provided at least one follow-up assessment 6 to 12 weeks into the program. We observed a significant reduction in depression scores at follow-up (mean –5.9, SD 5.2), with 74.5% of participants shifting downward in symptom severity by at least one diagnostic category. A similar pattern of reduction was observed for anxiety scores among participants who had moderate anxiety (mean –6.1, SD 5.4). Supplementary ITT analysis that included the entire cohort of 1512 participants with baseline data or last-value carryforward showed a slightly attenuated but still significant reduction in PHQ-8 scores at follow-up.

Examining the relationship between program usage and improvement in depression symptoms revealed that participants who more extensively used features across the platform had a modestly greater likelihood of reliable improvement in depression scores at follow-up (OR 1.3, 95% CI 1.1-1.5). This limited dose-response relationship was seen in the context of highly right-skewed usage data overall.

Drawn largely from a medically complex adult population, cardiometabolic conditions, particularly diabetes and obesity, were highly prevalent in this population. Given the very high burden of metabolic disease in this population and the availability of interventions for these conditions on the Vida platform, although the study was not specifically designed for this, we wanted to explore these relationships in order to prepare for future research.

Our preliminary analyses showed significantly greater reduction in body weight among participants who had a reliable improvement in depression symptoms compared to those who did not show symptom improvement at follow-up. As Figure 3 illustrates, this trend was consistent throughout the program tenure and culminated in a mean body weight change of 4.5% (SD 6.3; reliable improvement in PHQ-8 scores) versus 1.5% (SD 6.1; no reliable improvement in PHQ-8 scores) at month 9 of the program.

As with much of the literature, it is impossible to infer causation from these results, but it opens a hypothesis for future research that there may be opportunities for synergy in the treatment of co-occurring mental and metabolic disorders [37,62]. Given the massive and overlapping burden of these diseases in the US population, if such synergy could be realized through efficient, digitally delivered interventions, it would certainly be welcome.

Comparison With Prior Work

The principal finding of improvements in depressive symptoms is consistent with the broader body of research on digitally delivered CBT programs for the treatment of mild to moderate depression and anxiety [9]. This finding reinforces, in a larger population, similar results previously published from the Vida CBT Program [23].

The relationship between program usage and outcomes in digital interventions, broadly, and DMHIs, in particular, remains a rich area of research. While this study was not designed to evaluate this, our results are consistent with several similar studies suggesting a limited dose-response phenomenon. That is, greater usage seems to correlate with improved outcomes but may not do so monotonically, as participants self-regulate their usage.
to their needs, needs that are difficult for a researcher to observe with any meaningful precision [18,63-66]. Throughout the literature, interrelations among treatments for comorbid mental and physical disorders have been noted before. The direction of causality is unclear and, indeed, may be expected to run in both directions; that is, depression and anxiety may contribute to obesity and diabetes just as much as the reverse is true. In some cases, other factors may be driving both [62,67-69]. Preliminary research in this vein has, like this study, shown some hint of synergistic treatment effects in traditional care settings [70-72].

Limitations
This study had several important limitations. The lack of a control group and the retrospective design prevents drawing any causal inferences. Missing follow-up data was also a challenge, with data unavailable for 45.0% of the overall cohort (Table 1). We attempted to mitigate this limitation with an ITT analysis using baseline data or last-value carryforward as appropriate. While this still showed a significant relationship, the improvement in depression scores was weaker when accounting for those participants.

The gender imbalance across the study cohort was notable, with 83.8% females overall and 86.0% in the treatment group. At the overall level, this is a well-characterized phenomenon of care use, particularly for mental health, in North America and Europe. Despite seemingly comparable prevalence of mental distress by gender, ratios of 2 women for every man seeking care are not unusual in the literature [73,74]. The US National Institute of Mental Health has even sponsored campaigns to address barriers to males seeking mental health care, and there is literature on this phenomenon’s underlying causes [56,75]. In this study, there was also a small but significant increase in the proportion of females in the treatment group. It may have been that whatever phenomenon inhibits males from seeking care, in general, also made males in this study less likely to follow up with treatment and assessment. This and the difference in age between the two groups could represent some self-selection of participants into the treatment group who were more likely to improve. While the ITT analysis attempts to adjust for this, some bias remains a possibility. More broadly, the limited uptake of the Vida intervention by males may limit its generalizability in a larger population.

This study was not specifically designed to investigate changes in body weight and glycemic control as related to improvements in depression and was underpowered to detect a change in HbA1c. Future research is certainly needed to explore these potential relationships. Causality in any such relationship, should it bear out, could take any possible form and would need careful study to dissect.

Conclusions
This study provides further evidence that CBT-based DMHIs can be effective tools in treating symptoms of depression and anxiety. Improvements by whole diagnostic categories were common in this cohort. Usage of the platform correlated with improved outcomes in a pattern consistent with participants self-regulating usage to their individual needs. Furthermore, this was seen in a population with a very high prevalence of co-occurring physical disorders. More research is required to determine if there may be opportunities for synergistic treatment of mental and physical disorders through a similar modality.

Acknowledgments
We are deeply indebted to the hard work of all the therapists, dietitians, and health coaches who supported these participants. We are grateful to Chris Mosunic for helpful comments on the manuscript.

Authors' Contributions
AV and MS conceived and designed the study. MM led the intervention design. AV, PR, and BF acquired, analyzed, and interpreted the data. AV and MS drafted, structured, and edited the manuscript. All authors discussed the results and contributed to the final manuscript.

Conflicts of Interest
All authors are current or former employees of Vida Health and hold share options.

References


Abbreviations

- **CBT**: cognitive behavioral therapy
- **DMHI**: digital mental health intervention
- **GAD-7**: 7-item Generalized Anxiety Disorder scale
- **HbA₁c**: hemoglobin A₁c
- **ITT**: intention-to-treat
- **OR**: odds ratio
- **PHQ-8**: 8-item Patient Health Questionnaire
- **RCI**: reliable change index

---

©Aarathi Venkatesan, Benjamin Forster, Prasanna Rao, Melissa Miller, Michael Scahill. Originally published in JMIR Formative Research (https://formative.jmir.org), 05.07.2022. 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.
Optimization of the Chronic Kidney Disease–Peritoneal Dialysis App to Improve Care for Patients on Peritoneal Dialysis in Northeast Thailand: User-Centered Design Study

Eakalak Lukkanalikitkul¹, MD; Sawinee Kongpetch¹, MD; Wijittra Chotmongkol¹, MD; Michael G Morley², MSc, MD; Siririt Anutrakulchai³, MD, PhD; Chavis Srichan⁴, PhD; Bandit Thinkhamrop⁵, PhD; Theenatchar Chunghom⁶, RN; Pongsai Wiangnon⁶, RN; Wilaiphorn Thinkhamrop⁵, MPH, DrPH; Katharine E Morley⁷, MPH, MD

¹Center of Excellence in Kidney Diseases, Srinagarind Hospital, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand
²Ophthalmic Consultants of Boston, Harvard Medical School, Boston, MA, United States
³Division of Nephrology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand
⁴Department of Computer Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen, Thailand
⁵Data Management and Statistical Analysis Center, Faculty of Public Health, Khon Kaen University, Khon Kaen, Thailand
⁶Kidney Unit, Srinagarind Hospital, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand
⁷Massachusetts General Hospital Center for Global Health, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Katharine E Morley, MPH, MD
Massachusetts General Hospital Center for Global Health
Harvard Medical School
125 Nashua St. Suite 722
Boston, MA, 02114
United States
Phone: 1 617 726 2000
Email: kemorley@mgh.harvard.edu

Abstract

Background: The prevalence of peritoneal dialysis (PD) in Thailand is increasing rapidly in part because of Thailand’s Peritoneal Dialysis First policy. PD is a home-based renal replacement therapy in which patients with chronic kidney disease perform up to 4 exchanges of dialysate fluid per day in the peritoneal cavity. Overhydration is one of the most common complications in patients on PD and is associated with increased morbidity and mortality. To monitor hydration status, patients collect hydration metrics, including body weight, blood pressure, urine output, and ultrafiltration volume, from each dialysis cycle and enter this information into a PD logbook. The chronic kidney disease-PD (CKD-PD) app with near-field communication (NFC) and optical character recognition (OCR) was developed to automate hydration metric collection. The information was displayed in the app for self-monitoring and uploaded to a database for real-time monitoring by the PD clinic staff. Early detection and treatment of overhydration could potentially reduce the morbidity and mortality related to overhydration.

Objective: This study aims to identify usability issues and technology adoption barriers for the CKD-PD app with NFC and OCR and a monitoring system and to use this information to make rapid cycle improvements.

Methods: A multidisciplinary team of nephrologists, PD clinic nurses, computer programmers, and engineers trained and observed 2 groups of 5 participants in the use of the CKD-PD app with NFC and OCR and a monitoring system. The participants were observed using technology in their homes in 3 phases. The data collected included the Unified Theory of Acceptance and Use of Technology questionnaire, think-aloud observation, user ratings, completion of hydration metrics, and upload of hydration metrics to the central database. These results were used by the team between phases to improve the functionality and usefulness of the app.

Results: The CKD-PD app with NFC and OCR and a monitoring system underwent 3 rapid improvement cycles. Issues were identified regarding the usability of the NFC and OCR data collection, app stability, user interface, hydration metric calculation, and display. NFC and OCR improved hydration metric capture; however, issues remained with their usability. App stability and user interface issues were corrected, and hydration metrics were successfully uploaded by the end of phase 3. Participants’ scores...
on technology adoption decreased but were still high, and there was enthusiasm for the self-monitoring and clinical communication features.

**Conclusions:** Our rapid cycle process improvement methodology identified and resolved key barriers and usability issues for the CKD-PD app with NFC and OCR and a monitoring system. We believe that this methodology can be accomplished with limited training in data collection, statistical analysis, and funding.

*(JMIR Form Res 2022;6(7):e37291) doi:10.2196/37291*

**KEYWORDS**

peritoneal; dialysis; peritoneum; mobile health; mHealth; rapid cycle process improvement methodology; home monitoring; near-field communication; monitor; kidney; rapid cycle improvement; quality improvement; process improvement; methodology; nephrology; nephrologist; internal medicine; computer program; Unified Theory of Acceptance and Use of Technology; UTAUT; usability; interface; metric capture; barrier; renal; mobile phone

**Introduction**

**Chronic Kidney Disease and Peritoneal Dialysis in Northeast Thailand**

Chronic kidney disease (CKD) is a major health problem in Thailand because of its high prevalence, cost of treatment, significant morbidity and mortality, and substantial impact on the quality of life of patients and their families. From 2017 to 2018, the Chronic Kidney Disease Prevention in the Northeast of Thailand (CKDNET) project found that the prevalence of CKD was 27% in the rural provinces of northeast Thailand [1], primarily because of rising rates of diabetes, hypertension, and other primary renal diseases [2]. Despite this increased need, there is only 1 nephrologist for every 593,000 population compared with every 44,000 population in Bangkok [3]. In 2008, the Thai government adopted the *Peritoneal Dialysis First* policy for renal replacement therapy under its universal health care coverage scheme, increasing access in low-resource settings, with approximately 21% of patients on peritoneal dialysis (PD) in northeast Thailand [4]. Patients on PD manage their PD at home manually or by using a PD cycler to deliver dialysate fluid through a catheter placed in the peritoneal cavity, where the fluid remains for several hours. They enter hydration metric data, including body weight, blood pressure, urine output, and ultrafiltration volume for each cycle, in handwritten notebooks for review by a nephrologist at bimonthly clinic appointments.

**User Design and Evaluation of Successful Adoption of the CKD-PD App**

In 2018, the Data Management and Statistical Analysis Center, Faculty of Public Health, Khon Kaen University, developed the CKD-PD mobile app to help nephrologists and patients on PD manage fluid status to prevent overhydration. It is a common complication in patients on PD. Overhydration increases morbidity because of PD-related peritonitis, stroke, congestive heart failure, major adverse cardiac events, and mortality [4-9]. The CKD-PD app has been of interest not only to nephrologists but also to the Thai Health Security Office as an intervention for improving the care of patients on PD. Early treatment of overhydration can decrease these related complications, thereby reducing hospitalization and health care expenses [10].

Many apps fail because of a lack of evaluation and removal of barriers to user adoption, fidelity of the technology, and design of the health care delivery system in which the app will be deployed [11,12]. Achieving rapid design and deployment of digital health interventions are challenges facing the successful adoption and implementation of mobile health (mHealth) technologies [13]. This is especially true in low- and middle-income countries where mHealth interventions have been touted as solutions to a wide variety of health care challenges [14-16]; however, little is known about how they perform [17-19]. Before studying the effectiveness of the CKD-PD in managing overhydration in a real-world setting, a user design study was conducted using rapid cycle process improvement methods.

The objectives of this study were to (1) optimize the design and usability of the CKD-PD app and test the app in the context in which it is deployed using rapid cycle process improvement methods; (2) evaluate automatic data entry features using near-field communication (NFC) and optical character recognition (OCR) technology; and (3) identify and address the practical challenges that influence the successful adoption of the CKD-PD app and remote monitoring system in a real-world, low-resource setting.

**Methods**

**Study Population**

The study was conducted at Srinagarind Hospital, Khon Kaen University, between November 1, 2020, and April 30, 2021. Patients on PD for >3 months without a change in their PD prescription were invited to participate if they met the following inclusion criteria: aged ≥18 years, having access to a smartphone capable of running the CKD-PD app, and willing to allow research staff to observe their use of the CKD-PD app with NFC and OCR features and monitoring equipment in their home. Vulnerable populations specified as pregnant women, children, prisoners, individuals who were institutionalized, or those unable to participate in home data collection were excluded. Informed consent was obtained by the trained research staff. A total of 10 participants were enrolled and divided into group 1 (participants 1-5) and group 2 (participants 6-10). Baseline demographic data, including age, sex, education level, time on PD, and whether they used continuous ambulatory PD (CAPD) or automated PD (APD), were collected at enrollment. Participants were given prepaid cards to cover the cost of study related to internet or cellular data expenses.
The CKD-PD app is available for free download in Android and iOS formats and is designed for daily hydration metric collection for patients on PD (Multimedia Appendix 1). Users can enter hydration metrics using manual input or voice recognition. The CKD-PD app graphically displays a patient’s hydration metrics over time and uploads them to the CKDNET database stored in the Thai Care Cloud data repository [20], which is accessible to the PD clinic staff. Nephrologists can set individual hydration parameters for a patient so that they can self-monitor their hydration status. Alerts can be set to notify patients on PD and PD clinic staff when there are actionable hydration metric abnormalities, triggering a prompt review by a nephrologist and allowing the early detection and treatment of overhydration. Another key feature of the CKD-PD app is a direct link using LINE, a social messaging app widely used in Thailand, between the patient and PD clinic, facilitating communication about symptoms and changes in PD management.

A new prototype using NFC and OCR to automate the entry of hydration metric data into the CKD-PD app from measurement devices was developed. NFC uses radio frequency communication and wirelessly transmits data to an NFC-enabled device when it is placed within 4 cm. This offers a simple, low-cost solution that does not require an external power source or pairing, similar to Bluetooth. It is commonly used for mobile payments ad transit cards and, more recently, with subdermal glucose sensors [21-25]. NFC radio frequency tags with unique ID numbers can be added to medical devices such as scales. Users tap a card with an NFC receiver tag on the NFC-equipped scale and then tap it on the NFC card reader, and the data are transferred to the CKD-PD (Multimedia Appendix 2). The CKD-PD app is also equipped with OCR technology, which uses a smartphone camera to capture the digital output from the blood pressure machine and store it in the CKD-PD app.

### Study Design

#### Overview

There were 3 user design phases separated by 2 improvement cycles. During the improvement cycle, the app improvement team analyzed the results from the previous phase and revised the CKD-PD app using the NFC, OCR, and monitoring system (Table 1). The app improvement team included study nephrologists, computer engineers and app developers, and the Data Management and Statistical Analysis Center Thai Care Cloud database team. The app improvement team met as needed to define technology and user design issues, develop solutions, and test modifications.

### Table 1. Overview and timeline of research activities.

<table>
<thead>
<tr>
<th>Research activity</th>
<th>Phase 1 (group 1)</th>
<th>IC³ 1</th>
<th>Phase 2 (group 2)</th>
<th>IC 2</th>
<th>Phase 3 (groups 1 and 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week</td>
<td>0</td>
<td>1-2</td>
<td>3</td>
<td>4</td>
<td>5-8</td>
</tr>
<tr>
<td>In-clinic training and observation</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>UTAUTb survey</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Home observation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Completion of hydration metrics</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Contact with PDc clinic</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Validation of hydration metrics</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Observation logbook</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

³IC: improvement cycle.

bUTAUT: Unified Theory of Acceptance and Use of Technology.

cPD: peritoneal dialysis.

#### Phase 1

#### Overview

Group 1 participants received training on the use of the CKD-PD app with the NFC and OCR and monitoring system during week 0. In-home observations were conducted once during weeks 1 to 4, followed by research activities 2 to 3 (see the following sections) during weeks 1 and 2. During weeks 3 and 4, the participants used the CKD-PD app with the NFC and OCR system at their homes and completed research activities 4 to 6 (see in the following sections).

#### Improvement Cycle 1

After completion of phase 1, the results were summarized and presented with the completion of hydration metrics and clinical contact results to the app improvement team. The app improvement team analyzed the results and modified the components of the system, including measurement devices, NFC, OCR, app design, and programming. The user processes were also adjusted.

#### Phase 2

#### Overview

Group 2 participants received training in the use of the CKD-PD app with the NFC and OCR and monitoring system during week...
8. In-home observations were conducted during weeks 9 to 12 using the modified CKD-PD app with the NFC and OCR and monitoring system. Research activities 2 to 3 were conducted during weeks 9 and 10, followed by research activities 4 to 6 in weeks 11 and 12.

Improvement Cycle 2
After the completion of phase 2, the app improvement team reviewed the new results from participant observations, hydration metric completion, and clinic contacts. Additional issues were identified, and new solutions were developed and tested. Phase 3 was launched after the completion of the improvement cycle 2 modifications.

Phase 3
Overview
Participants from groups 1 and 2 performed research activities 2 to 6 again during weeks 17 to 20 using the modified CKD-PD app and monitoring system modified without NFC and OCR. This allowed for the evaluation of the usability and functionality of the CKD-PD app without NFC and OCR. In phase 3, participants were asked to rate the entry of body weight and dialysate volumes with NFC again based on their earlier experience for comparison with phases 1 and 2. In addition, research activity 7 (see the following sections) was performed to compare using the usual practice of entering their hydration metrics by hand into a logbook with the CKD-PD app.

Final Improvements
After completion of phase 3, the app improvement team made final improvements based on the results of the phase 3 home observation and hydration metric collection using the CKD-PD app and monitoring system without NFC and OCR.

Description of Study Activities
The details of the study activities are described in the following sections.

In-Clinic Training
Participants were trained to use the CKD-PD app with NFC and OCR and home monitoring equipment in the PD clinic by research staff and PD nurses, starting with the CKD-PD app downloaded onto the participant’s mobile phone and account registration. Research assistants demonstrated the use of the CKD-PD app, followed by instructions on how to use the NFC and home monitoring equipment, how the uploaded hydration metrics can be monitored by the PD clinic, and how they can self-monitor at home on the app. A set of NFC hydration metric data collection equipment (body weight scale, NFC card, NFC card reader and connectors, and blood pressure machine) was prepared for each participant to use at home. They were trained to set up the equipment using a teaching video (Multimedia Appendix 2) followed by hands-on practice with a research assistant. During the study period, the participants were instructed to use the NFC data collection system and equipment, in addition to the standard method of recording their hydration metrics in a logbook.

Unified Theory of Acceptance and Use of Technology Survey
A structured interview questionnaire based on the Unified Theory of Acceptance and Use of Technology (UTAUT) model was completed at the beginning and end of each phase to collect perceptions of user technology acceptance and usability [26]. The questionnaire has 6 domains, each with 3 questions representing different factors affecting technology use and adoption. It was translated into the Thai language, and each question was scored on a 5-point Likert scale, with 1=strongly disagree and 5=strongly agree. The final score for each domain is the sum of the scores of the 3 questions for that domain and ranges from 3=strongly disagree to 15=strongly agree. The questionnaire was explained by the research staff and self-administered by the participants (Multimedia Appendix 3).

Home Observation
Research staff conducted the home observation of participants using the CKD-PD app with NFC and OCR using an observation guide (Multimedia Appendix 4), using the think-aloud method [27]. They asked the participants what they liked and disliked about each task and feature. Participants were asked to rate each feature as follows: 1=good, 2=neutral, and 3=not good. This information was recorded by research assistants using handwritten notes and summarized for use by the app improvement team.

Completion of Hydration Metrics
During the 2-week home use period, the number of times the participant successfully uploaded each of the required hydration metric values (body weight, blood pressure, and use of the CKD-PD app) was recorded. Participants using CAPD required 4 dialysate exchange volumes per day, whereas those using APD only required 1 each day; thus, the number of required hydration metrics varied among participants. Successful completion of hydration metrics included the entry of hydration metrics into the app and accurate upload to the CKDNET database.

Contact With PD Clinic
The number of times the participants contacted the PD clinic during all 3 phases was collected, along with the reason for contact.

Validation of Hydration Metrics
The hydration metrics collected by each participant were validated by comparing the values uploaded to the CKDNET database with the results recorded in their logbooks during the study period. Participants sent screenshots from the app, or the research staff reviewed the data on the participant’s smartphone to confirm the correct data entry in the CKD-PD app. Successful validation of hydration metrics was defined by the completion of data entry by the participant and agreement between values recorded in their logbook and the uploaded results.

Observation of Logbook Use
Participants were observed while entering hydration metrics in their logbooks. They were asked what they liked and disliked
about each task and feature and rated each feature as 1 = good, 2 = neutral, and 3 = not good.

**Ethics Approval**

This study was approved by the Ethics Committee for Human Research, Faculty of Medicine, Khon Kaen University, Thailand (project number HE621494), and the Mass General Brigham institutional review board (protocol number 2019P002648). All participants provided written informed consent in the Thai language.

**Results**

**Participant Characteristics**

Phase 1 participants characteristics had a mean age of 46 years (SD 10.3) and a mean time on PD of 5.5 years (SD 3.8); were female (2/5 40%); used the PD method of CAPD (3/5, 60%) or APD (2/5, 40%); and had educational backgrounds of high school (2/5, 40%), bachelor’s degree (1/5, 20%), and postbachelor’s degree (2/5, 40%). Phase 2 participant characteristics had a mean age of 48 years (SD 15.7) and a mean time on PD of 1.3 years (SD 2.5); were female (3/5, 60%); used the PD method of CAPD (1/5, 20%) or APD (4/5, 80%); and had educational backgrounds of high school (2/5, 40%), bachelor’s degree (2/5, 40%), and postbachelor’s degree (1/5, 20%). One of the group 2 participants did not complete phase 3 because of sudden death.

In comparison, the characteristics of patients on PD at Srinagarind Hospital were as follows: mean age 49 years (SD 13); mean time on PD 3 years (SD 29.7); female (33/74, 45%); PD method CAPD (42/74, 57%) or APD (32/74, 43%); and educational background of primary school (37/74, 50%), high school (12/74, 16%), bachelor’s degree (23/74, 31%), and postbachelor’s degree (2/74, 3%).

**UTAUT Survey**

The UTAUT survey scores for each participant are presented in Table 2 as the total score (sum of all 6 domains) and the scores for each domain. The difference in the total score between the beginning of phase 1 or phase 2 and the end of phase 3 is also presented. Detailed results, including the scores at the beginning and end of all phases, are available in Multimedia Appendix 5.

**Table 2.** Difference in Unified Theory of Acceptance and Use of Technology total scoresa by participant and domain between the beginning of phase 1 or 2 and end of phase 3 (detailed results in Multimedia Appendix 5).

<table>
<thead>
<tr>
<th>Participant numberb</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Values, mean</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 1 or 2</td>
<td>89</td>
<td>63</td>
<td>86</td>
<td>81</td>
<td>87</td>
<td>79</td>
<td>66</td>
<td>75</td>
<td>71</td>
<td>77</td>
<td>77</td>
</tr>
<tr>
<td>Phase 3</td>
<td>74</td>
<td>76</td>
<td>80</td>
<td>74</td>
<td>77</td>
<td>72</td>
<td>80</td>
<td>79</td>
<td>72</td>
<td>76</td>
<td>N/A</td>
</tr>
<tr>
<td>Difference</td>
<td>−15</td>
<td>13</td>
<td>−6</td>
<td>−7</td>
<td>−10</td>
<td>−7</td>
<td>N/A</td>
<td>5</td>
<td>8</td>
<td>−5</td>
<td>−2.7</td>
</tr>
<tr>
<td><strong>Difference by domains</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance expectancy</td>
<td>−2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>−1</td>
<td>−1</td>
<td>N/A</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.33</td>
</tr>
<tr>
<td>Effort expectancy</td>
<td>−2</td>
<td>4</td>
<td>1</td>
<td>−1</td>
<td>1</td>
<td>−1</td>
<td>N/A</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.33</td>
</tr>
<tr>
<td>Social influence</td>
<td>−3</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>−2</td>
<td>−1</td>
<td>N/A</td>
<td>−1</td>
<td>1</td>
<td>−1</td>
<td>−0.67</td>
</tr>
<tr>
<td>Voluntariness</td>
<td>−3</td>
<td>−3</td>
<td>−5</td>
<td>−2</td>
<td>−3</td>
<td>−3</td>
<td>N/A</td>
<td>2</td>
<td>3</td>
<td>−4</td>
<td>−2</td>
</tr>
<tr>
<td>Intention to use</td>
<td>−2</td>
<td>3</td>
<td>−4</td>
<td>−4</td>
<td>−6</td>
<td>0</td>
<td>N/A</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>−0.9</td>
</tr>
<tr>
<td>Facilitating conditions</td>
<td>−3</td>
<td>3</td>
<td>0</td>
<td>−1</td>
<td>1</td>
<td>−1</td>
<td>N/A</td>
<td>1</td>
<td>0</td>
<td>−1</td>
<td>−0.1</td>
</tr>
</tbody>
</table>

aScores for individual questions from 1 (strongly disagree) to 5 (strongly agree). The domain scores ranged from 5 to 15. The total score for all the domains ranged from 30 to 90.
bParticipants 1 to 5: phase 1 and phase 3; participants 6 to 10: phase 2 and phase 3.
cN/A: not applicable; participant 7 expired before phase 3.

Of the 9 participants who completed phase 1 or 2 and phase 3, 6 (67%) individuals had a decrease in the total UTAUT score, ranging from −5 to −15 points. There were 33% (3/9) of individuals who had an increase in the total UTAUT score, ranging from 5 to 13 points. The mean score at the beginning of phases 1 or 2 was 77 (SD 8.8), with a range of 63 to 89. The mean score at the end of phase 3 was 76 (SD 3.2), with a range of 72 to 80. In group 1, the domains of voluntariness (−3) and intention to use (−2.6) showed the largest decrease in mean difference between the beginning of phase 1 and the end of phase 3, with all participants reporting a negative score in voluntariness and 80% (4/5) of participants reporting a negative score in intention to use. In group 2, there were small decreases in voluntariness (−0.5), social influence (−0.5), and facilitating conditions (−0.25) and an increase in intention to use (1.25) between phases 2 and 3 (Multimedia Appendix 5).

**Participant Observation**

The home observation results for each participant using the CKD-PD app with the NFC and OCR and monitoring system during the home observations are summarized by phase for all participants in Table 3, with detailed responses in Multimedia Appendix 6. Feature ratings are provided for each phase. The participants rated almost all features as *good*, except for some...
data entry tasks using NFC, which were rated as neutral. Participants liked automatic data entry with NFC; however, technical issues such as slow transfer of time and difficulty tapping the card resulted in lower scores. Participants were asked to rate the entry of body weight and dialysate volume using NFC at the start of phase 3 based on their experience in phases 1 or 2 for comparison with manual data entry used in phase 3. The average rating for data entry with NFC was 2.3 for each task compared with 1.1 for manual data entry (Multimedia Appendix 6). Scores for all tasks using the handwritten logbook were nearly all higher than the CKD app with or without NFC and OCR, consistent with the preference for using the app.

### Table 3. Feature and task ratings[^a] from participant observation in phases 1, 2, and 3 (detailed results in Multimedia Appendix 6).

<table>
<thead>
<tr>
<th>Feature</th>
<th>Phase 1[^b], mean (range)</th>
<th>Phase 2[^c], mean (range)</th>
<th>Phase 3[^d], mean (range)</th>
<th>Phase 3[^e], mean (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open screens</td>
<td>1.1 (1-2)</td>
<td>1 (1)</td>
<td>1.6 (1-3)</td>
<td>N/A[^f]</td>
</tr>
<tr>
<td>Enter body weight</td>
<td>1.1 (1-2)</td>
<td>1.2 (1-2)</td>
<td>1.1 (1-2)</td>
<td>1.4 (1-2)</td>
</tr>
<tr>
<td>Enter blood pressure</td>
<td>1.1 (1-2)</td>
<td>1.6 (1-3)</td>
<td>1.1 (1-2)</td>
<td>1.4 (1-3)</td>
</tr>
<tr>
<td>Enter dialysate</td>
<td>1.1 (1-2)</td>
<td>1.2 (1-2)</td>
<td>1.1 (1-2)</td>
<td>1.8 (1-3)</td>
</tr>
<tr>
<td>View metrics</td>
<td>1.3 (1-2)</td>
<td>1 (1)</td>
<td>1.1 (1-2)</td>
<td>1.7 (1-3)</td>
</tr>
<tr>
<td>Interpret metrics</td>
<td>1.1 (1-2)</td>
<td>1.2 (1-2)</td>
<td>1.1 (1-2)</td>
<td>N/A</td>
</tr>
<tr>
<td>Communications</td>
<td>1 (1)</td>
<td>1 (1)</td>
<td>1 (1)</td>
<td>1.2 (1-3)</td>
</tr>
<tr>
<td>User incentives</td>
<td>1.1 (1-2)</td>
<td>1.2 (1-2)</td>
<td>1.2 (1-3)</td>
<td>1.9 (1-3)</td>
</tr>
</tbody>
</table>

[^a]: Rating of task or feature: 1=good, 2=neutral, and 3=not good.
[^b]: Group 1 using the chronic kidney disease–peritoneal dialysis app with near-field communication or optical character recognition data entry; participants 1 to 5.
[^c]: Group 2 using the chronic kidney disease–peritoneal dialysis app with near-field communication or optical character recognition data entry; participants 6 to 10.
[^d]: Group 1 and 2 using the chronic kidney disease–peritoneal dialysis app with manual entry; participants 1 to 10, excluding 7.
[^e]: Group 1 and 2 using logbook; participants 1 to 10, excluding 7.
[^f]: N/A: not applicable.

In general, participants found that manual data entry was easier to perform than using the NFC and OCR system. The use of NFC and OCR presented multiple challenges. Some of these issues were solved by providing tools such as an extension device for tapping the card to eliminate the need to bend over and touch the card to the scale on the floor and training participants to wait long enough for data transfer. Some issues were more difficult to troubleshoot within the time frame of this study, such as (1) lack of space in participants’ homes for NFC setup and access to power outlets; (2) internet instability resulting in slow data transfer; (3) size and design of the NFC-assembled weight scale, making it difficult for older or obese patients to stand on; and (4) lack of an alert to indicate that the NFC device was ready to place the NFC card for reading. Most patients also did not like to use the OCR function to enter the blood pressure readings because of poor image clarity and variations in ambient home lighting conditions.

Several important issues related to the functionality of the CKD-PD app were detected and corrected. The dialysate in and out volumes were reported by PD cycle, and the net daily ultrafiltration volume was not accurately calculated and displayed. There were issues with the display of hydration metrics, such as incorrect scale on the graph and a lack of previous results for comparison. During home observation of participants using the CKD-PD app with the NFC and OCR and monitoring system, participants expressed concerns about the usability of the app. An example was slowness when opening each icon, which was determined to be from opening multiple apps at once, and slow cellular or Wi-Fi network speeds were contributing factors. BMI was misinterpreted as overweight because of translation issues between the English and Thai languages. This was corrected by changing the wording so it would not be confused with weight. Participant height had to be entered daily to calculate the BMI. As this measurement did not change, the app was modified so that it automatically entered the participant height obtained daily. The original font was small and difficult to read. This was adjusted, and the readability of the screens was improved.

### Completion and Validation of Hydration Metrics

The percentage completion of the hydration metrics for each participant was collected for all 3 phases (Table 4). The mean percentages of completion in phases 1 and 2 were 88% (SD 19) and 83% (SD 6.3), respectively, with a range of 57% to 100%. This was compared with the mean percentage of completion in phase 3 for all participants (68%, SD 18.6, range 18%-100%; Multimedia Appendix 7). The most common reasons for not collecting the hydration metrics in phases 1 and 2 were not knowing how to enter the data, forgetting to enter the data, problems with the system uploading the data to the CKDNET database, and switching from Android to iOS systems. The reasons for incomplete hydration metric collection in phase 3 (not using NFC) were forgetting to send the data to the CKDNET and switching from Android to iOS.
Table 4. Percentage completion of hydration metrics (detailed in Multimedia Appendix 7).

<table>
<thead>
<tr>
<th>Participant</th>
<th>Phase 1a</th>
<th>Phase 2b</th>
<th>Phase 3c</th>
<th>Difference (%)d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total, N</td>
<td>Done, n (%)</td>
<td>Total, N</td>
<td>Done, n (%)</td>
</tr>
<tr>
<td>1</td>
<td>84</td>
<td>80 (95)</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>2</td>
<td>66</td>
<td>39 (59)</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>3</td>
<td>42</td>
<td>42 (100)</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>4</td>
<td>84</td>
<td>73 (87)</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>5</td>
<td>48</td>
<td>48 (100)</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>6</td>
<td>—</td>
<td>—</td>
<td>42</td>
<td>38 (91)</td>
</tr>
<tr>
<td>7f</td>
<td>—</td>
<td>—</td>
<td>39</td>
<td>35 (90)</td>
</tr>
<tr>
<td>8</td>
<td>—</td>
<td>—</td>
<td>42</td>
<td>37 (88)</td>
</tr>
<tr>
<td>9</td>
<td>—</td>
<td>—</td>
<td>30</td>
<td>24 (57)</td>
</tr>
<tr>
<td>10</td>
<td>—</td>
<td>—</td>
<td>80</td>
<td>27 (90)</td>
</tr>
<tr>
<td>Values, mean</td>
<td>65</td>
<td>56 (88)</td>
<td>46.6</td>
<td>32 (83)</td>
</tr>
</tbody>
</table>

aPhase 1: participants 1 to 5.
bPhase 2: participants 6 to 10.
cPhase 3: participants 1 to 10.
dDifference in percentage completion between phase 1 or 2 and phase 3.
eNot available.
fParticipant 7 expired before completion of the study.

Of the 9 participants who entered hydration metric data in phase 1 or 2 and phase 3, there were 4 participants who had a greater than 29 percentage point decrease in their hydration metric completion. The other 5 participants had minimally decreased or increased hydration metric completion, ranging from –11% to 9%.

Validation of the hydration metrics improved from phases 1 or 2 and phase 3 (Table 5). In phases 1 and 2, the main issues were the upload of inaccurate body weight and ultrafiltration metric results to the CKDNET. This was because of problems with the NFC weight scale and app calculation of the ultrafiltration volume. In phase 3, these issues were resolved, the hydration metrics were fully validated with participants correctly entering the data in the CKD-PD app, and the results entered were accurately uploaded to the CKDNET.
Table 5. Validation of hydration metrics.

<table>
<thead>
<tr>
<th>Participant</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Values, n (%)[^{a}]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1[^{b}]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body weight</td>
<td>No[^{c}]</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>_[^{d}]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>Yes[^{e}]</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>4 (80)</td>
</tr>
<tr>
<td>Dialysate in</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>3 (60)</td>
</tr>
<tr>
<td>Dialysate out</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>3 (60)</td>
</tr>
<tr>
<td>Ultrafiltration volume</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>2 (40)</td>
</tr>
<tr>
<td>Phase 2[^{f}]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body weight</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>3 (60)</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>5 (100)</td>
</tr>
<tr>
<td>Dialysate in</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>5 (100)</td>
</tr>
<tr>
<td>Dialysate out</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>4 (80)</td>
</tr>
<tr>
<td>Ultrafiltration volume</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>3 (60)</td>
</tr>
<tr>
<td>Phase 3[^{g}]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body weight</td>
<td>Yes</td>
<td>ND[^{h}]</td>
<td>ND</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A[^{i}]</td>
<td>Yes</td>
<td>Yes</td>
<td>7 (100)</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>9 (100)</td>
</tr>
<tr>
<td>Dialysate in</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>9 (100)</td>
</tr>
<tr>
<td>Dialysate out</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>9 (100)</td>
</tr>
<tr>
<td>Ultrafiltration volume</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A</td>
<td>Yes</td>
<td>ND</td>
<td>9 (100)</td>
</tr>
</tbody>
</table>

\[^{a}\]Number and percentage of hydration metric values entered and uploaded with the correct results to Chronic Kidney Disease Prevention in the Northeast of Thailand.

\[^{b}\]Phase 1: participants 1 to 5.

\[^{c}\]Not uploaded accurately to Chronic Kidney Disease Prevention in the Northeast of Thailand.

\[^{d}\]Not available.

\[^{e}\]Uploaded accurately to Chronic Kidney Disease Prevention in the Northeast of Thailand.

\[^{f}\]Phase 2: participants 6 to 10.

\[^{g}\]Phase 3: participants 1 to 10.

\[^{h}\]ND: not done by participant.

\[^{i}\]N/A: not applicable; participant 7 expired before completion of the study.

Clinic Contacts
The number of times each participant contacted the PD clinic during each phase is presented in Table 6. All contacts were made using the Line messaging app by chat, audio, or video calls, except for 1 in-person contact during phase 1. In phase 1, there were 104 clinic contacts compared with 13 and 22 in phases 2 and 3, respectively, and most were related to issues using the CKD-PD app. In phase 3, 59% (13/22) of the contacts were related to clinical concerns.

Table 6. Number and reason for contacting the peritoneal dialysis clinic by phase.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Total number of contacts, N</th>
<th>Clinical issue, n (%)</th>
<th>App issue, n (%)</th>
<th>Other, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1</td>
<td>104</td>
<td>0 (0)</td>
<td>99 (95.2)</td>
<td>5 (4.8)</td>
</tr>
<tr>
<td>Phase 2</td>
<td>13</td>
<td>3 (23.1)</td>
<td>10 (76.9)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Phase 3</td>
<td>22</td>
<td>13 (59.1)</td>
<td>9 (40.9)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

Improvement Team Activities
After the completion of each phase, the research team summarized the findings for the app improvement team (Multimedia Appendix 8). In phase 1, troubleshooting issues with NFC and OCR were the focus of the improvement activities, along with app design issues. In phase 2, additional work was performed with the NFC data entry; however, more focus was given to resolving issues with how the hydration metric data were displayed and uploaded to the CKDNET database. In phase 3, the NFC prototype was not used as the participants preferred to input data manually into the CKD-PD
app, and the upload of the hydration metric data to the CKDNET database was not reliable using NFC. The app improvement team made final improvements to improve the stability of the data transfer from the CKD-PD app to the CKDNET database, added the total daily ultrafiltration volume to the CKDNET, and set parameters for overhydration alerts.

Discussion

Principal Findings

Overview

mHealth apps are touted as having great potential to transform the delivery of health care; however, the real-world development of mHealth apps in low- and middle-income countries reveals few success stories [12,17]. Often, a good app idea fails to succeed as the design does not meet the users’ needs and is difficult to use. Our research describes how a rapid cycle process improvement strategy helped researchers understand the benefits and challenges of using an mHealth app for use by patients on PD in a low-resource setting and optimize its features. The process revealed valuable insights into the factors influencing user attitudes, identifying technological and design flaws, and addressing barriers to user adoption of the CKD-PD app and monitoring system. Moreover, we believe that this methodology can be accomplished without requiring significant training in data collection, statistical analysis, or funding. The general principles of the user design process are applicable to a wide variety of locations, contexts, and subject domains.

User Adoption

The total scores for all the participants for most domains were in the agree and strongly agree categories, demonstrating a strong willingness to adopt the technology at the beginning of the study. Over the course of the study, 67% (6/9) of participants who completed both assigned phases demonstrated decreased interest in using the technology, although the total scores remained in the agree range or above. The decreases in the total score were primarily driven by decreases in the domains of voluntariness and intention to use. We believe this is likely because of issues with the NFC and OCR data entry features and app speed and functionality. We acknowledge that the UTAUT questionnaire is not typically used to evaluate technology adoption over time; however, we found that it provided insights into how our participants’ views changed over the course of using the CKD-PD app and monitoring system, although they had limited generalizability. The small sample size limited the statistical validity of our findings. However, we found the information useful as a sentiment analysis regarding user experience and provided insights into how different individuals may adopt the CKD-PD app.

CKD-PD App Issues

Design issues centered on simplicity and ease of use compared with a pencil and paper logbook. Simplifying the use of the app, for example, opening the app and the organization of the screens, readability of the fonts, input of the data, and graphical representation of trends, clearly needed to be refined based on user feedback. PD requires a substantial commitment of time to collect and record data from their families throughout the day, whether in a logbook or an app. Our participants generally reacted positively to the self-monitoring features and quickly engaged with this new functionality but became frustrated when some features did not work well for them.

Technology Issues

Technology issues included problems with the NFC devices, unstable internet access, slow internet speed, app stability, uploading data to the CKD-NET cloud, and software issues causing it to freeze or not upload data properly. The participant observation revealed that NFC and OCR were good solutions for automatic data entry but that they need to be seamlessly integrated into the measurement devices; for example, the location of NFC on the scales on the floor was inconvenient, slow data transfer, and poor quality image capture with OCR. Despite these barriers, participants liked the concept, and with further development, we believe that NFC and OCR could be simple and effective methods of mitigating the barrier of data entry [28,29], as has been done with other devices such as glucose monitors.

Internet access, especially in low-income rural areas, is frequently slow or unstable, making the use of digital apps cumbersome. Some patients were concerned about the cost of sending or receiving data, although internet access was subsidized for the participants in this study. These will be challenging issues to remedy; however, if successful in reducing complications from overhydration, it may prove cost-effective in the long run to provide financial support for internet access for patients on PD, given the alternative costs of not using the app.

Programming issues resulting in app instability and data upload errors were critical problems that were identified. These were fixable by our software engineers and resulted in improved validation of the hydration metrics in phase 3. The completion of the hydration metrics dropped between phases 1 and 3. We believe this occurred as participants in phases 1 and 2 used NFC, which automatically uploaded the hydration metrics, and decreased after the team elected not to use NFC and OCR in phase 3 because of usability issues and inaccurate upload of the hydrate metrics to the CKDNET with NFC and OCR. The hydration metric accuracy improved with this change and resulted in 100% validation of the hydration metrics in phase 3. However, when manual entry was used in phase 3, participants had to remember to upload their data, resulting in incomplete hydration metrics. In the final improvements, a reminder alert addressed this issue.

Communication Issues

Contact with the PD clinic was much more frequent in phase 1 than in phase 2, although the participants in both phases were new users of the app. The study staff felt that this was because of differences in individual comfort when using the CKD-PD app. In phase 3, there were fewer contacts about the app, although there were 10 participants compared with 5 in phases 1 and 2. The contacts in phase 3 were about clinical concerns, suggesting increased confidence in self-monitoring features. In general, participants gave high ratings and positive comments.
about the ease of communication with the PD clinic using the CKD-PD app.

**Impact on Future Deployment and Development of CKD-PD**

Our research findings provided critically useful information for the optimization of the CKD-PD app and monitoring system. We plan to use it in a randomized controlled trial to further evaluate the efficacy of CKD-PD in the early detection and treatment of overhydration.

We have developed training materials for CKD-PD app users based on the insights gained from this study and plan to share them with future CKD-PD app users. We have improved awareness of user adoption issues and realize that this will be an iterative process as more patients on PD use the app. We believe that with additional engineering and design work, NFC is a potential solution for automating data entry, and we plan to pursue the further development of these features.

**Limitations**

Our study has several limitations. The NFC system and CKD-PD app were initially tested together; however, the NFC issues were too complicated to solve within the time frame and budget of the study period. Through targeted observations and data collection, we were able to tease out NFC and data entry issues from the app design and cloud-based monitoring processes.

We identified potential sources of bias. Observation bias could have influenced our results as participant observations were conducted by research staff from the PD clinic. Patients on PD, cared for by nephrologists and PD nurses from their clinic, may be inclined to report feedback they thought the PD staff wanted to hear. Confirmation bias may also be a concern as research staff may be looking for patterns or use issues that confirm opinions they already hold about the CKD-PD app, NFC and OCR data entry, and the monitoring system. Our sample size was small; however, this is consistent with recommendations for user design studies, indicating that a few participants can uncover most usability issues [30]. We also acknowledge the potential for selection bias in our participants, given their higher educational level than the average population in rural northeast Thailand, although the participants were representative of the Srinagarind Hospital PD patient population. We selected participants who had baseline comfort with smartphones for the user design study, although it limits generalizability as they were more interested and engaged in the design process.

An additional weakness is that we did not conduct an in-depth investigation into why our participants responded to the UTAUT questionnaire as much as they did. Although it has been validated and used in Thai populations [26,31], we did not consider the impact of age, gender, education, and cultural influences on UTAUT scores [32,33]. We acknowledge that UTAUT results have limited generalizability.

**Comparison With Prior Work**

Our research included several recommended strategies for mHealth app development and evaluation [34,35]. Our iterative rapid cycle process improvement approach used a multidisciplinary team, including patients, computer engineers and programmers, nephrologists, and PD nurses, to evaluate the CKD-PD app. We conducted our study activities in the participants’ homes and PD clinics where it would be used [36,37]. Our structured observation using the think-aloud method provided the app improvement team with a broad sense of what worked and what did not work, in addition to specific actionable feedback. Finally, our study participants were selected from the same study population as those who will use the CKD-PD app in northeast Thailand.

Our research team found several critical usability and functionality issues during the study. This experience reinforces the importance of these steps before an mHealth app can be successfully implemented [12,13,17]. Although our current NFC and OCR data entry features are not optimized, our results support the idea that app users want these features [27].

**Conclusions**

In conclusion, using rapid cycle progress improvement methodologies with a multidisciplinary team proved to be a useful strategy for optimizing the CKD-PD app and monitoring system. Our experience aligns with insights and recommendations regarding mHealth technology design and evaluation that using a multidisciplinary team in the context of the system in which it is used is essential for successful deployment. This process revealed critical issues that once addressed, position the CKD-PD app and monitoring system to achieve its potential to improve health outcomes.

**Acknowledgments**

The research reported in this publication was supported by the Fogarty International Center of the National Institutes of Health under award number R21TW010963. The content is the sole responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors would like to acknowledge the Chronic Kidney Disease Northeast Thailand research project staff for their assistance in organizing and assisting with home observations. The authors would like to thank Ms Thankyalak Rattanasawad; the staff of the Center of Excellence in Kidney Diseases, Srinagarind Hospital, Faculty of Medicine at Khon Kaen University; Dr Jessica Haberer and the Massachusetts General Hospital Center for Global Health; Dr Nathaphop Chaichaya, Data Management and Statistical Analysis Center, Faculty of Public Health, Khon Kaen University Thailand; the other staff from the Data Management and Statistical Analysis Center and the Faculty of Public Health at Khon Kaen University; and Pobporn Danvirutai from T. Robotics, Co Ltd, Bangkok, Thailand.

https://formative.jmir.org/2022/7/e37291

JMIRES2022 | vol. 6 | iss. 7 | e37291 | p.46

(page number not for citation purposes)
Authors’ Contributions
SA, EL, KEM, MGM, and BT designed the study, including working with the Data Management and Statistical Analysis Center app development team. CS developed and oversaw the use of near-field communication or optical character recognition and integration with the chronic kidney disease–peritoneal dialysis app. EL, SK, WC, TC, and PW collected and analyzed the data and drafted the manuscript. SA, KEM, and MGM assisted with the data analysis and critically revised the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Chronic Kidney Disease–Peritoneal Dialysis app screenshots.
[PDF File (Adobe PDF File), 2613 KB - formative_v6i7e37291_app1.pdf ]

Multimedia Appendix 2
Using Chronic Kidney Disease–Peritoneal Dialysis app (near-field communication version).
[MP4 File (MP4 Video), 15806 KB - formative_v6i7e37291_app2.mp4 ]

Multimedia Appendix 3
Unified Theory of Acceptance and Use of Technology questionnaire on adoption and use of the Chronic Kidney Disease–Peritoneal Dialysis app.
[PDF File (Adobe PDF File), 95 KB - formative_v6i7e37291_app3.pdf ]

Multimedia Appendix 4
Participant observation guide.
[PDF File (Adobe PDF File), 42 KB - formative_v6i7e37291_app4.pdf ]

Multimedia Appendix 5
Unified Theory of Acceptance and Use of Technology scores by participant and domain between phases 1 and 3 or phases 2 and 3.
[PDF File (Adobe PDF File), 174 KB - formative_v6i7e37291_app5.pdf ]

Multimedia Appendix 6
Summary of participant observation phases 1, 2, and 3.
[PDF File (Adobe PDF File), 74 KB - formative_v6i7e37291_app6.pdf ]

Multimedia Appendix 7
Detailed hydration metric completion.
[PDF File (Adobe PDF File), 78 KB - formative_v6i7e37291_app7.pdf ]

Multimedia Appendix 8
Recommendations from the app improvement team.
[PDF File (Adobe PDF File), 55 KB - formative_v6i7e37291_app8.pdf ]

References


20. DAMASAC Data Management and Statistical Analysis Center. Faculty of Public Health, Khon Kaen University. URL: [https://www.thaiarcercle.org/] [accessed 2022-02-10]


Abbreviations

APD: automatic peritoneal dialysis
CAPD: continuous ambulatory peritoneal dialysis
CKD: chronic kidney disease
CKDNET: Chronic Kidney Disease Prevention in the Northeast of Thailand
CKD-PD: chronic kidney disease–peritoneal dialysis
mHealth: mobile health
NFC: near-field communication
OCR: optical character recognition
PD: peritoneal dialysis
UTAUT: Unified Theory of Acceptance and Use of Technology

https://formative.jmir.org/2022/7/e37291

Please cite as:
URL: https://formative.jmir.org/2022/7/e37291
doi:10.2196/37291
PMID:35793137
Experimental Implementation of NSER Mobile App for Efficient Real-Time Sharing of Prehospital Patient Information With Emergency Departments: Interrupted Time-Series Analysis

Kiyomitsu Fukaguchi1*, MD; Tadahiro Goto2*, MD, MPH, PhD; Tadatsugu Yamamoto1*, MD; Hiroshi Yamagami1*, MD

1Department of Emergency Medicine, Shonan Kamakura General Hospital, Kamakura-shi, Kanagawa, Japan
2TXP Research, TXP Medical Co Ltd, Tokyo, Japan
*all authors contributed equally

Corresponding Author:
Kiyomitsu Fukaguchi, MD
Department of Emergency Medicine
Shonan Kamakura General Hospital
1370-1
Okamoto
Kamakura-shi, Kanagawa, 247-0072
Japan
Phone: 81 467 46 1717
Fax: 81 467 45 0190
Email: fukaskgh@gmail.com

Abstract

Background: With the aging society, the number of emergency transportations has been growing. Although it is important that a patient be immediately transported to an appropriate hospital for proper management, accurate diagnosis in the prehospital setting is challenging. However, at present, patient information is mainly communicated by telephone, which has a potential risk of communication errors such as mishearing. Sharing correct and detailed prehospital information with emergency departments (EDs) should facilitate optimal patient care and resource use. Therefore, the implementation of an app that provides on-site, real-time information to emergency physicians could be useful for early preparation, intervention, and effective use of medical and human resources.

Objective: In this paper, we aimed to examine whether the implementation of a mobile app for emergency medical service (EMS) would improve patient outcomes and reduce transportation time as well as communication time by phone (ie, phone-communication time).

Methods: We performed an interrupted time-series analysis (ITSA) on the data from a tertiary care hospital in Japan from July 2021 to October 2021 (8 weeks before and 8 weeks after the implementation period). We included all patients transported by EMS. Using the mobile app, EMS can send information on patient demographics, vital signs, medications, and photos of the scene to the ED. The outcome measure was inpatient mortality and transportation time, as well as phone-communication time, which was the time for EMS to negotiate with ED staffs for transport requests.

Results: During the study period, 1966 emergency transportations were made (n=1033, 53% patients during the preimplementation period and n=933, 47% patients after the implementation period). The ITSA did not reveal a significant decrease in patient mortality and transportation time before and after the implementation. However, the ITSA revealed a significant decrease in mean phone-communication time between pre- and postimplementation periods (from 216 to 171 seconds; −45 seconds; 95% CI −71 to −18 seconds). From the pre- to postimplementation period, the mean transportation time from EMS request to ED arrival decreased by 0.29 minutes (from 36.1 minutes to 35.9 minutes; 95% CI −2.20 to 1.60 minutes), without change in time trends. We also introduced cases where the app allowed EMS to share accurate and detailed prehospital information with the emergency department, resulting in timely intervention and reducing the burden on the ED.

Conclusions: The implementation of a mobile app for EMS was associated with reduced phone-communication time by 45 seconds (22%) without increasing mortality or overall transportation time despite the implementation of new methods in the real
The city has a high ratio of the aging population, with 53,517 people and an average age of 48.8 years as of 2020. Kamakura City is a city in Kanagawa Prefecture, Japan, with a population of approximately 2000 based on the number of patients transported in the past.

Introduction

With the aging society, the number of emergency transportation has been growing [1]. Researchers used a nationwide database in Japan and reported that the annual emergency transportations increased from 4 million in 2000 to 6 million in 2020 [1]. The substantial increase in emergency transportation causes depletion of medical resources and emergency department (ED) overcrowding, resulting in poor patient outcomes [2,3]. Indeed, in Japan, approximately 10,000 emergency patients were turned away by at least 6 hospitals annually [4].

When a patient has stroke, acute myocardial infarction, or severe trauma, it is important to transport them to an appropriate hospital in order to provide appropriate management immediately [5-9]. Although accurate diagnosis in the prehospital setting is challenging, real-time sharing of detailed patient information should facilitate appropriate transportation and management [10,11]. However, at present, patient information is mainly communicated by telephone, which has the potential risk of communication errors such as mishearing. In addition, photos of the patient, patient monitors (eg, electrocardiogram), and the accident scene can contain important information that emergency physicians would want to know in advance. Therefore, the implementation of an app that provides on-site, real-time information to emergency physicians could be useful for early preparation, intervention, and effective use of medical and human resources [5-8,10-12].

Thus, in this study, we examined whether the implementation of a mobile app for emergency medical services (EMS) could improve patient outcomes. We hypothesized that patient mortality would be improved by shortening the transportation time and time for EMS to negotiate with ED staffs for transport requests.

Methods

Study Design and Settings

We performed an interrupted time-series analysis to examine the change in outcomes before and after the introduction of a mobile app for EMS. This study was conducted using data on EMS transportations by Kamakura City Fire Department to Shonan Kamakura General Hospital, a tertiary care hospital in Japan, from July 8, 2021, to October 27, 2021. Kamakura City is a city in Kanagawa Prefecture, Japan, with a population of 172,948 people and an average age of 48.8 years as of 2020. The city has a high ratio of the aging population, with 53,517 people (31.1%) aged ≥65 years. Kamakura City has 8 fire departments that make 29,9 emergency transportations per day and 10,896 per year [13]. The Emergency Department of Shonan Kamakura General Hospital had 43,506 ED visits by patients in 2020. Among these patients, 14,925 (34.3%) were transported to the emergency department by EMS, which was the highest annual number of emergency transportations of all hospitals in Japan that year. All patients transported by EMS were accepted [14]. The number of patients transported by the Kamakura City EMS during the study period (16 weeks) was estimated to be approximately 2000 based on the number of patients transported in the past.

Ethics Approval

This study protocol was approved by the Ethics Committee of Shonan Kamakura General Hospital (approval number: TGE01663-024; approval date: February 25, 2021). The need for informed consent was waived due to the retrospective nature of the study. For the publication of images and personal information, the patients’ consent was obtained, and their consent was recorded in their medical charts. The institutional review board approved the use of patient information for this study.

Apps Used and Their Features

We implemented an app for EMS named NEXT Stage ER mobile (NSER mobile) on September 2, 2021. The NSER mobile is an app that was developed to reduce the complexity of information entry for EMS teams and to increase the efficiency of information sharing with EDs. The app is equipped with high-precision voice input and optical character recognition functions that enable timely information sharing by sending information to hospitals in the original text (eg, medications and past medical history) and in coded form (eg, International Statistical Classification of Diseases and Related Health Problems, 10th revision codes). We provide a description of the usage process of the NSER mobile by EMS and hospitals in Multimedia Appendices 1 and 2. All EMS staffs took 3-hour lectures to learn to use the app before implementation.

The developer of NSER mobile is TXP Medical Co Ltd. A free trial for this product is available through the website of TXP Medical [15]. The system’s accuracy and validity are now being examined in a multicenter study (UMIN-CTR ID: UMIN000045775), but it is fundamentally the same as the NEXT Stage ER system, which is a well-validated ED information system [16]. In addition, in this study, we did not use table data from the app. All data are abstracted from the Kamakura City Fire Department and patient registry of ED patients of Shonan Kamakura General Hospital. Thus, the only data we extracted from the NEXT Stage ER mobile system for
this study are images (though more detailed data exist in the cloud server of the system).

**Study Population**

All patients transported to Shonan Kamakura General Hospital by the Kamakura City EMS during the study period were included in the study. Exclusion criteria were cases with missing data on call times between the hospital and EMS and EMS activity times (26/1992, 1.3%).

**Outcome Measures**

The primary outcome was inpatient mortality. Secondary outcomes were the overall transportation time (the time from the patient’s call for an ambulance to arrival at the hospital), and phone-communication time (the time for EMS to negotiate with ED staffs for transport requests).

**Statistical Analysis**

Cases with missing data on the time spent in EMS and the emergency department (ED) were excluded from the analysis (26/1992, 1.3%). Using data that met the inclusion criteria, we performed an interrupted time-series analysis with a linear regression model to examine whether the implementation of the mobile app had an impact on the outcomes. An interrupted time-series analysis is a quasi-experimental design for evaluating the effectiveness of population-level health interventions implemented at a clearly defined point in time and is thus widely used to evaluate the effectiveness of interventions [17-19]. The time unit was weeks (ie, 8 time points before and 8 time points after implementation).

Additionally, we described specific cases where the immediate sharing of patient information and photos with NSER mobile led to rapid diagnosis and intervention.

All statistical analyses were conducted using R version 4.1.1 (R Foundation for Statistical Computing).

**Results**

Among the 1992 patients transported by EMS, we excluded 26 patients (1.3%) with missing data on the transport time or the phone communication time, and the remaining 1966 (98.7%) patients were eligible for this study. The mean age was 66.7 years (SD 25.4), and 49% (n=963) were men. Of the 1966 patients, 1033 (53%) were transported during the preimplementation period, and 933 (47%) were transported in the postimplementation period. The patient characteristics in the pre- and postimplementation periods were similar. Additionally, there were no significant differences in the region of transportation and the EMS team numbers (Table 1).
Table 1. Characteristics of patients transported by emergency medical services.

<table>
<thead>
<tr>
<th>Patient characteristics and variables</th>
<th>Patients transported using the app (n=1033)</th>
<th>Patients transported in the usual way (n=933)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>65.8 (27.2)</td>
<td>68.1 (23.5)</td>
<td>.13</td>
</tr>
<tr>
<td><strong>Age profile (years), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤18</td>
<td>98 (9.5)</td>
<td>48 (5.1)</td>
<td></td>
</tr>
<tr>
<td>18-64</td>
<td>307 (29.7)</td>
<td>232 (24.9)</td>
<td></td>
</tr>
<tr>
<td>65-84</td>
<td>324 (31.4)</td>
<td>343 (36.8)</td>
<td></td>
</tr>
<tr>
<td>≥85</td>
<td>304 (29.4)</td>
<td>310 (33.2)</td>
<td></td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td>.80</td>
</tr>
<tr>
<td>Male</td>
<td>508 (49.2)</td>
<td>464 (49.7)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>525 (50.8)</td>
<td>469 (50.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Number of calls to the hospital, mean (SD)</strong></td>
<td>1.07 (0.28)</td>
<td>1.05 (0.25)</td>
<td>.05</td>
</tr>
<tr>
<td><strong>Region of emergency medical services, n (%)</strong></td>
<td></td>
<td></td>
<td>.17</td>
</tr>
<tr>
<td>Kamakura</td>
<td>162 (15.7)</td>
<td>147 (15.8)</td>
<td></td>
</tr>
<tr>
<td>Ofuna</td>
<td>176 (17.0)</td>
<td>181 (19.4)</td>
<td></td>
</tr>
<tr>
<td>Fukasawa</td>
<td>203 (19.7)</td>
<td>155 (16.6)</td>
<td></td>
</tr>
<tr>
<td>Tamanawa</td>
<td>139 (13.4)</td>
<td>128 (13.7)</td>
<td></td>
</tr>
<tr>
<td>Koshigoe</td>
<td>121 (11.7)</td>
<td>109 (11.7)</td>
<td></td>
</tr>
<tr>
<td>Imaizumi</td>
<td>74 (7.2)</td>
<td>82 (8.8)</td>
<td></td>
</tr>
<tr>
<td>Shitirigahama</td>
<td>91 (8.8)</td>
<td>60 (6.4)</td>
<td></td>
</tr>
<tr>
<td>Zyoumyouzi</td>
<td>67 (6.5)</td>
<td>71 (7.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Level of consciousness (JCS⁹), n (%)</strong></td>
<td></td>
<td></td>
<td>.70</td>
</tr>
<tr>
<td>0</td>
<td>524 (50.7)</td>
<td>443 (47.5)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>163 (15.8)</td>
<td>178 (19.1)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>143 (13.8)</td>
<td>123 (13.2)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>90 (8.7)</td>
<td>79 (8.5)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>47 (4.5)</td>
<td>45 (4.8)</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>14 (1.4)</td>
<td>15 (1.6)</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>5 (0.5)</td>
<td>4 (0.4)</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>7 (0.7)</td>
<td>8 (0.9)</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>7 (0.7)</td>
<td>4 (0.4)</td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>33 (2.9)</td>
<td>34 (3.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Vital signs, mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systolic blood pressure (mmHg)</td>
<td>133 (45.9)</td>
<td>135 (50.6)</td>
<td>.35</td>
</tr>
<tr>
<td>Diastolic blood pressure (mmHg)</td>
<td>79.4 (27.8)</td>
<td>79.6 (30.7)</td>
<td>.91</td>
</tr>
<tr>
<td>Pulse rate (per min)</td>
<td>89.1 (29.3)</td>
<td>84.5 (27.2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Respiratory rate (per min)</td>
<td>21.2 (6.33)</td>
<td>20.8 (5.67)</td>
<td>.08</td>
</tr>
<tr>
<td>Saturation (%)</td>
<td>89.2 (24.5)</td>
<td>87.9 (26.8)</td>
<td>.25</td>
</tr>
<tr>
<td>Body temperature (°C)</td>
<td>35.0 (8.43)</td>
<td>34.4 (9.04)</td>
<td>.16</td>
</tr>
<tr>
<td><strong>Classification of diseases⁹, n (%)</strong></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Endogenous disease</td>
<td>748 (72.4)</td>
<td>680 (72.9)</td>
<td></td>
</tr>
<tr>
<td>Trauma</td>
<td>270 (26.1)</td>
<td>234 (25.1)</td>
<td></td>
</tr>
<tr>
<td>Cardiac arrest</td>
<td>15 (1.5)</td>
<td>19 (2.0)</td>
<td></td>
</tr>
</tbody>
</table>
### Severity of illness at the ED\textsuperscript{a}, n (%)

<table>
<thead>
<tr>
<th>Illness Level</th>
<th>Preimplementation</th>
<th>Postimplementation</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minor illness</td>
<td>375 (36.3)</td>
<td>351 (37.6)</td>
<td>.70</td>
</tr>
<tr>
<td>Moderate illness</td>
<td>567 (54.9)</td>
<td>492 (52.7)</td>
<td></td>
</tr>
<tr>
<td>Serious illness</td>
<td>76 (7.3)</td>
<td>71 (7.6)</td>
<td></td>
</tr>
<tr>
<td>Death</td>
<td>15 (1.5)</td>
<td>19 (2.0)</td>
<td></td>
</tr>
</tbody>
</table>

### Disposition at the ED, n (%)

<table>
<thead>
<tr>
<th>Disposition</th>
<th>Preimplementation</th>
<th>Postimplementation</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discharge</td>
<td>657 (63.6)</td>
<td>549 (58.8)</td>
<td>.04</td>
</tr>
<tr>
<td>Admission</td>
<td>308 (29.8)</td>
<td>295 (31.6)</td>
<td></td>
</tr>
<tr>
<td>Transfer to another hospital for admission</td>
<td>50 (4.8)</td>
<td>68 (7.3)</td>
<td></td>
</tr>
<tr>
<td>Death</td>
<td>18 (1.7)</td>
<td>21 (2.3)</td>
<td></td>
</tr>
</tbody>
</table>

### Prognosis, n (%)

<table>
<thead>
<tr>
<th>Prognosis</th>
<th>Preimplementation</th>
<th>Postimplementation</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-hospital death</td>
<td>53 (5.1)</td>
<td>46 (4.9)</td>
<td>.56</td>
</tr>
</tbody>
</table>

\textsuperscript{a}JCS: Japan Coma Scale.

\textsuperscript{b}The severity of illness at the emergency department is classified as follows: minor illness—patient can return home after treatment; moderate illness—patient requires inpatient treatment, but the disease severity is low and can be managed in a general ward; serious illness—multiorgan failures such as respiratory or circulatory failure requiring monitoring, a ventilator, vasopressors such as catecholamines, and admission to an intensive care unit; death—cardiac arrest on arrival at the hospital.

\textsuperscript{c}ED: emergency department.

### Inpatient Mortality

Of the 1966 eligible patients, 53 (5.1\%) died in hospital in the preimplementation period and 46 (4.9\%) died in hospital in the postimplementation period. From the pre- to postimplementation period, the proportions of in-hospital deaths among patients who were transported to EDs during each period decreased by 5\% (95\% CI \textasciitilde11\% to 1\%), followed by a decreasing trend relative to preimplementation of \textasciitilde1\% per week (95\% CI \textasciitilde2\% to 1\%). There was no significant change in inpatient mortality. Figure 1 shows the proportions of in-hospital deaths among patients who were transported to emergency departments during each period. On September 2, 2021, EMS began transportation using the new app in place of the traditional method. The proportion of in-hospital deaths among the transported patients is plotted for 8-week periods before and after the implementation. For in-hospital mortality, the $R^2$ for the preimplementation model was 0.26, while the $R^2$ for the postimplementation model was 0.23. The results of interrupted time series analysis on inpatient mortality are shown in Table 2.
Table 2. Results of interrupted time-series analysis on inpatient mortality.

<table>
<thead>
<tr>
<th>Time-series analysis</th>
<th>Estimate (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trends in inpatient mortality before implementation</td>
<td>0.01 (0.00 to 0.02)</td>
<td>.07</td>
</tr>
<tr>
<td>Absolute change in the inpatient mortality before and after implementation</td>
<td>−0.05 (−0.11 to 0.01)</td>
<td>.11</td>
</tr>
<tr>
<td>Trends in inpatient mortality after implementation</td>
<td>0.00 (−0.01 to 0.01)</td>
<td>.50</td>
</tr>
<tr>
<td>Change in slope before and after implementation</td>
<td>−0.01 (−0.02 to 0.01)</td>
<td>.40</td>
</tr>
</tbody>
</table>

The Transportation Time From EMS Request to ED Arrival

The mean transportation time from EMS request to ED arrival was 35.9 minutes (SD 9.7 minutes) in the preimplementation period and 36.1 minutes (SD 8.5 minutes) in the postimplementation period. From the pre- to postimplementation period, the mean transportation time from EMS request to ED arrival decreased by 0.29 minutes (95% CI −2.20 to 1.60 minutes), followed by a decreasing trend relative to preimplementation of −0.33 minutes per week (95% CI −0.74 to 0.07; Table 3). Figure 2 shows the mean transportation time from the emergency call to arrival at the hospital. On September 2, 2021, EMS began transporting using the new app in place of the traditional method. The mean transportation time is plotted for 8-week periods before and after the implementation. For transportation time, the $R^2$ for the preimplementation model was 0.30, and the $R^2$ for the postimplementation model was 0.28.

Table 3. Results of interrupted time-series analysis on transportation time from emergency medical services to emergency department arrival.

<table>
<thead>
<tr>
<th>Time-series analysis</th>
<th>Estimate (min), 95% CI</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trends in mean transportation time before implementation</td>
<td>0.23 (−0.06 to 0.51)</td>
<td>.11</td>
</tr>
<tr>
<td>Absolute change in the transportation time before and after implementation</td>
<td>−0.29 (−2.20 to 1.60)</td>
<td>.70</td>
</tr>
<tr>
<td>Trends in mean transportation time after implementation</td>
<td>−0.10 (−0.39 to 0.18)</td>
<td>.40</td>
</tr>
<tr>
<td>Change in slope before and after implementation</td>
<td>−0.33 (−0.74 to 0.07)</td>
<td>.10</td>
</tr>
</tbody>
</table>
Phone-Communication Time Between EMS Teams and Hospital

The mean time of phone communication between EMS and ED staffs was 216 (SD 107) seconds in the preimplementation period and 171 (SD 120) seconds in the postimplementation period. From the pre- to postimplementation period, the phone-communication time decreased by 45 seconds (95% CI −71 to −18 seconds), followed by an increasing trend relative to preimplementation of +2.9 seconds per week (95% CI −2.7 to 8.6; Table 4). The mean phone-communication time between EMS and the hospital is shown in Figure 3. On September 2, 2021, the EMS started using the app for transportation in place of the conventional method. The mean phone-call time is plotted for 8-week periods before and after the implementation. For phone-communication time, the $R^2$ for the preimplementation model was 0.78, while the $R^2$ for the postimplementation model was 0.72.

Table 4. Results of interrupted time-series analysis on phone-communication time.

<table>
<thead>
<tr>
<th>Time-series analysis</th>
<th>Estimate (s), 95% CI</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trends in mean phone-communication time before implementation</td>
<td>−0.44 (−4.4 to 3.6)</td>
<td>.80</td>
</tr>
<tr>
<td>Absolute change in the phone-communication time before and after implementation</td>
<td>−45.0 (−71.0 to −18.4)</td>
<td>.003</td>
</tr>
<tr>
<td>Trends in mean phone-communication time after implementation</td>
<td>2.5 (−1.5 to 6.5)</td>
<td>.20</td>
</tr>
<tr>
<td>Change in slope before and after implementation</td>
<td>2.9 (−2.7 to 8.6)</td>
<td>.30</td>
</tr>
</tbody>
</table>
Specific Cases

We experienced several cases where the immediate sharing of patient information and photos using the NSER mobile has led to rapid diagnosis and intervention. For example, a 47-year-old man with a history of diabetes mellitus, who had visited the hospital the day before for chest pain, called for EMS again complaining of persistent chest pain. Upon EMS’s arrival, his electrocardiography monitor showed ST-segment elevation in lead II (Multimedia Appendix 3). Before arriving at the hospital, he was diagnosed with ST-segment elevation acute myocardial infarction. Upon consulting with the cardiologist, he was able to start percutaneous coronary intervention within ten minutes after the patient’s arrival at the ED. Multimedia Appendix 3 shows the monitor screen in the ambulance transporting a 47-year-old man who called EMS complaining of chest pain. Lead II showed ST-segment elevation, and the images were transmitted from EMS to the ED staff, leading to early diagnosis of acute myocardial infarction and percutaneous coronary intervention within 10 minutes of arrival at the hospital.

In another example, a 35-year-old man was riding a bicycle when he collided with a car traveling at 60 kilometers per hour. He had no memory when he was injured. While the mechanism of injury suggests a highly serious one, real-time information sharing (initial assessment of the patient, vital signs, photos of the injury scene, and damage to the bicycle) allowed us to determine that the injury was minor, with only a contusion on the left lower leg and the scalp. Consequently, we were able to reduce unnecessary preparation for the initial treatment, including surgery for damage control (Multimedia Appendices 4-6).

Discussion

Principal Findings

In this ITSA study, there was no substantial decrease in in-hospital mortality and overall transportation time between pre- and postimplementation periods. On the other hand, the implementation reduced phone-communication time by 22% without increasing mortality or overall transportation time despite the implementation of new methods in the real clinical setting.

The conventional communication between EMS and ED staff by telephone only poses substantial stress on EMS staff, exemplified by difficulty hearing and misidentification of information. Data related to patient demographics, vital signs, past medical history, and medications are likely misclassified due to such technical difficulties, and real-time information sharing could reduce such miscommunication. In addition, the app’s feature to share visual information via optical character recognition in a timely manner was useful to ensure that ED staffs are fully prepared to receive patients. For ED staff, timely monitoring and understanding of the situation offered by the EMS were useful for maintaining a high quality of clinical practice.

Comparison With Prior Work

Appropriate use of medical apps could lead to a seamless transition of management from prehospital to post-ED arrival. As reported in this study, apps can be used to obtain information such as medical history and prehospital electrocardiograms in advance. From such information, physicians can prepare for urgent interventions (eg, catheterization) before the patient arrives at the hospital. There have been several reports on the
usefulness of apps that provide prehospital information in emergency medicine. A system called ORION (Osaka Emergency Information Research Intelligent Operation Network system), introduced in Osaka City in 2013, reduced the number of cases that are difficult to transport [12]. Another study reported that the communication-type medical apps can be accurately used remotely, and information can be shared with the stroke team to prepare for rapid treatment [5,20,21].

The NSER mobile app is a digitalization tool for EMS in the clinical setting, and there are no patients for whom the app cannot be used. Nevertheless, we think that the system is more suitable for patients who need emergency interventions (eg, cardiac catheterization) [2-5,7] rather than those with cardiac arrest (EMS may not have enough time to use the app). Due to the limited sample size, we could not analyze data after stratifying by these variables. Thus, we are conducting further study in different settings to examine the effectiveness of the app.

**Strengths and Limitations**

This study has several strengths, a few of which are as follows. First, there were no similar studies on prehospital information transfer apps aimed at improving the efficiency of emergency patient transport without assuming a specific disease. Second, the interrupted time-series analysis estimates the effect of intervention on a population and is a study design without a control group. Third, the advantage of reducing phone-communication time through real-time information sharing is noticeable especially when the EMS is consulting multiple hospitals to accept patients at the same time, not to mention that in many cases it is difficult to transport patients, particularly in urban areas in Japan. Indeed, according to statistics from the Ministry of Health, Labor, and Welfare in 2016, even for critically ill patients, there were 10,039 cases (2.3%) in which the number of consultations to medical institutions was ≥4 times and 22,104 cases (5.0%) in which the time spent on site was 30 minutes or longer [22]. In such cases, given the tough negotiation with hospitals, the reduction of phone-communication time while efficiently sharing prehospital information should reduce the burden on EMS. We believe that the findings from this study allow us to consider the substantial contribution and potential benefits of mobile apps to emergency medical care.

Our study has several limitations. First, although we performed a 3-hour lecture for using the app prior to implementation, users may not have been able to get accustomed to the app quickly enough in the clinical setting. Despite this, there was clear improvement in phone-communication time immediately after its implementation. Second, there was no control group in our ITSA design [23]. Nonetheless, our findings are likely robust, given that there were no interventions other than the implementation of the app that may have affected the outcome. While the COVID-19 pandemic may have affected the assumptions of the interrupted time series analysis, the implementation of personal protective equipment for EMS was initiated on March 2, 2020. Therefore, change in practice due to the pandemic may not have substantially affected the EMS during the study period (July 8, 2021, to October 27, 2021). Third, we did not have information on the time to intervention at the ED (eg, time to urgent catheterization) and ED overcrowding. Therefore, further studies are needed to examine the impact of the app on clinical practice. Fourth, in this study, we only evaluated the observed values for 8 weeks before and after the intervention. A recent simulation-based study on ITSA reported that 12 preintervention and 12 postintervention time points may be required for a moderate intervention effect sizes [24]. Lastly, there is limited generalizability of our findings because our study was a single-center, retrospective observational study in Japan with a small sample size. In addition, EMS systems are different across countries. The extrapolation of our findings to other settings should be done with caution, and therefore additional large-scale studies are warranted.

**Future Directions**

As shown in the 2 cases, the implementation of a mobile app for efficient real-time sharing of prehospital patient information has potential to reduce the time to intervention, resulting in better patient outcomes. In addition, in Japan, especially in the urban areas, there is the difficulty in determining the hospital for emergency patient [1]; however, for instance, the average number of hospitals that EMS phoned to transport patient was 1 in Kamakura city during this study period, so the decision to transport a patient did not take extra time. Therefore, the app may reduce the overall transportation time by decreasing the number of calling from EMS to hospitals.

**Conclusions**

The implementation of a mobile app for EMS reduced phone-communication time by 22% without increasing mortality or overall transportation time despite the implementation of new methods in the real clinical setting. Real-time patient information sharing, such as the transfer of monitor images and photos of the accident site, could facilitate optimal patient care and resource use.

**Acknowledgments**

This study was supported by the grant “2021 Survey and Research Project on First Aid” from Japan Emergency Medical Foundation. The study is still in progress, and the results presented in this study are preliminary.

**Conflicts of Interest**

TG is the Chief Scientific Officer of TXP Medical Co Ltd.

Multimedia Appendix 1

https://formative.jmir.org/2022/7/e37301

JMIIR Form Res 2022 | vol. 6 | iss. 7 | e37301 | p.59

(page number not for citation purposes)
Workflow of an emergency medical services team using NEXT Stage ER mobile.

Multimedia Appendix 2
An example of sharing trauma images, monitors, and clinical information from emergency medical services to the emergency department staff.

Multimedia Appendix 3
Monitor screen image of a case of a 47-year-old man complaining of chest pain.

Multimedia Appendix 4
The damaged bicycle image of a case of a 35-year-old man with trauma.

Multimedia Appendix 5
Left lower extremity injuries of a case of a 35-year-old man with trauma.

Multimedia Appendix 6
Monitor screen image of a case of a 35-year-old man with trauma.

References
before-and-after observational study in Osaka City, Japan. JMIR Mhealth Uhealth 2017 Sep 11;5(9):e134 [FREE Full text] [doi: 10.2196/mhealth.8296] [Medline: 28893725]


Abbreviations

ED: emergency department
EMS: emergency medical services
ITSA: Interrupted time-series analysis
NSER mobile: NEXT Stage ER mobile

©Kiyomitsu Fukaguchi, Tadahiro Goto, Tadatsugu Yamamoto, Hiroshi Yamagami. Originally published in JMIR Formative Research (https://formative.jmir.org), 06.07.2022. 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.
Chinese Version of the Mobile Health App Usability Questionnaire: Translation, Adaptation, and Validation Study

Yi Shan1, Prof Dr; Meng Ji2, PhD; Wenxiu Xie3, MPhil; Rongying Li4, BM; Xiaobo Qian5, BEng; Xiaomin Zhang6, PhD; Tianyong Hao5, Prof Dr

1School of Foreign Studies, Nantong University, Nantong, China
2School of Languages and Cultures, University of Sydney, Sydney, Australia
3Department of Computer Science, City University of Hong Kong, Hong Kong, China
4School of Artificial Intelligence, South China Normal University, Guangzhou, China
5School of Computer Science, South China Normal University, Guangzhou, China
6Department of Linguistics, Macquarie University, Sydney, Australia

Corresponding Author:
Yi Shan, Prof Dr
School of Foreign Studies
Nantong University
Number 9, Seyuan Road
Nantong, 226019
China
Phone: 86 15558121896
Email: victorsyhz@hotmail.com

Abstract

Background: The usability of mobile health (mHealth) apps needs to be effectively evaluated before they are officially approved to be used to deliver health interventions. To this end, the mHealth App Usability Questionnaire (MAUQ) has been designed and proved valid and reliable in assessing the usability of mHealth apps. However, this English questionnaire needs to be translated into other languages, adapted, and validated before being utilized to evaluate the usability of mHealth apps.

Objective: This study aims to improve, further adapt, and validate the Chinese version of the MAUQ (C-MAUQ; interactive for patients) on Left-handed Doctor, one of the most popular “reaching out to patients” interactive mHealth apps with chatbot function in China, to test the reliability and cross-cultural adaptability of the questionnaire.

Methods: The MAUQ (interactive for patients) has been translated into Chinese and validated for its reliability on Good Doctor, one of the most influential “reaching out to patients” mHealth apps without chatbot function in China. After asking for the researchers’ approval to use this Chinese version, we adjusted and further adapted the C-MAUQ by checking it against the original English version and improving its comprehensibility, readability, idiomaticity, and cross-cultural adaptability. Following a trial survey completed by 50 respondents on wenjuanxing, the most popular online questionnaire platform in China, the improved version of the C-MAUQ (I-C-MAUQ) was finally used to evaluate the usability of Left-handed Doctor through an online questionnaire survey (answered by 322 participants) on wenjuanxing, to test its internal consistency, reliability, and validity.

Results: The I-C-MAUQ still retained the 21 items and 3 dimensions of the original MAUQ: 8 items for usability and satisfaction, 6 items for system information arrangement, and 7 items for efficiency. The translation problems in the C-MAUQ, including (1) redundancy, (2) incompleteness, (3) misuse of parts of speech, (4) choice of inappropriate words, (5) incomprehensibility, and (6) cultural difference–induced improper translation, were improved. As shown in the analysis of data obtained through the online survey, the I-C-MAUQ had a better internal consistency (ie, the correlation coefficient between the score of each item and the total score of the questionnaire determined within the range of 0.861-0.938; P<.01), reliability (Cronbach α=.988), and validity (Kaiser–Meyer–Olkin=0.973), compared with the C-MAUQ. It was effectively used to test the usability of Left-handed Doctor, eliciting over 80% of informants’ positive attitudes toward this mHealth app.

Conclusions: The I-C-MAUQ is highly reliable and valid for Left-handed Doctor, and suitable for testing the usability of interactive mHealth apps used by patients in China. This finding further confirms the cross-cultural validity, reliability, and adaptability of the MAUQ. We identified certain factors influencing the perceived usability of mHealth apps, including users’ age, gender, education, profession, and possibly previous experience with mHealth apps and the chatbot function of such apps.
Most notably, we found a wider acceptance of this new technology among young Chinese female college students who were more engaged in the interaction with health care chatbots. The age-, gender-, and profession-induced preference for new digital health interventions in China aligns with the findings in other similar studies in America and Malaysia. This preference identifies areas for further research on the social, cultural, and gender adaptation of health technologies.

(JMIR Form Res 2022;6(7):e37933) doi:10.2196/37933

KEYWORDS
mHealth app; usability; Chinese version of MAUQ; improved translation; validity; stability; reliability; cross-cultural adaptability; mobile phone

Introduction

Background
Mobile health (mHealth) apps have been applied to deliver health interventions (e.g., health education, health monitoring, recommendations on treatments) to alleviate the overburdened health systems in many countries. These apps can perform versatile tasks, including health management, behavior intervention, health data collection, self-diagnosis, disease management, medication management, rehabilitation, and acting as patient portals [1,2], improving medication compliance, saving time in diagnosis and treatment, and reducing medical costs [3-6]. Given these wide applications and diverse advantages, these apps need to be assessed for hidden expenses, heavy data entry burden, and interest loss [7] to ensure accurate data analysis before being put into use [8].

To effectively evaluate the usability of mHealth apps, different questionnaires were designed [9], among which the most popular are the System Usability Scale (SUS) and the Post-Study System Usability Questionnaire (PSSUQ) [10,11]. Although used to reliably measure certain usability aspects of mobile apps, the SUS and the PSSUQ, among others, failed to provide tailored information on the factors unique to mobile apps [10,12]. Zhou et al [9] developed and validated the mHealth App Usability Questionnaire (MAUQ), which was solely designed for assessing the usability of mHealth apps, attesting its reliability and validity. The MAUQ [9] was exclusively developed to evaluate the usability of mHealth apps. It has 4 versions designed to assess interactive or standalone mHealth apps among patients or health care providers. It shows a strong internal consistency, evidenced by the Cronbach α coefficients of its 3 dimensions (.895 for ease of use and satisfaction, .829 for system information arrangement, and .900 for usefulness) and the overall Cronbach α of .914. The items in the 3 dimensions are rated on a 7-point Likert scale from 1 (extremely strongly agree) to 7 (extremely strongly disagree). The usability of an app can be determined by calculating the total points and determining the average points of the responses to all statements: the closer the average is to 1, the higher the usability of the app [9].

Two more recent studies translated and adapted the MAUQ into Chinese [13] and Malay [14], respectively, finding that the Chinese and Malay versions exhibited high reliability and validity similar to those of the original English version [13,14]. The Chinese version of the MAUQ (C-MAUQ; interactive for patients) was testified to be reliable and valid, with content validity index of 0.952, Cronbach α of .912, value of test-retest reliability of 0.896, and value of the split-half reliability of 0.701 [14]. The Malay version of the MAUQ (standalone for patients) was proved to be reliable for evaluating the usability of the mHealth apps (Cronbach α=.946) [13]. Considering the painstaking efforts and considerable time and cost investment involved in developing new questionnaires [14], Marzuki et al [12] strongly recommended that established, accessible, and reliable questionnaires should be adapted, validated, and recorded cross-linguistically.

Left-handed Doctor is one of the most popular “reaching out to patients” [15] interactive mHealth apps in China. It integrates artificial intelligence technologies, such as deep learning, big data processing, semantic understanding, and interactive medical dialog with medicine and is committed to using artificial intelligence technology to expand the supply of high-quality medical resources. The Left-handed Doctor open platform provides solutions, such as smart hospitals, diagnostic robots for consultation rooms, intelligent online consultation, intelligent postdiagnosis management, and artificial intelligence internet hospitals. In combination with different application scenarios, it provides high-quality medical services for all parties, empowering the health care industry. Although it is popular among many people in China, no studies have empirically tested its usability using the C-MAUQ.

Objective
Informed by the MAUQ and its culturally adapted versions, this study aimed to testify further the reliability, validity, and cross-cultural adaptability of the MAUQ for its suitability to the mHealth app usability test. This was achieved by applying the improved version of the C-MAUQ (I-C-MAUQ) to Left-handed Doctor, one of the most popular “reaching out to patients” interactive mHealth apps with chatbot function in China. Two facts warrant this study: (1) the Left-handed Doctor app is different from the Good Doctor app: the former is empowered with the chatbot function, while the latter is not, and we thought that this difference would influence users’ perceived usability of these apps; and (2) the informants differ from those in Mustafa et al [13] in terms of age, gender, education, and profession, and we believed that these differences would also impact users’ perceived usability of these apps.

Methods

Overview
This study used the C-MAUQ [15] but made some improvements. The study was conducted from February 18 to March 8, 2022.
Improvement of the C-MAUQ

We first obtained the approval of the researchers [15] to use the C-MAUQ. Afterward, 2 translators (YS and MJ) independently adjusted this version by checking it against the original English version and improving its readability and idiomaticity. The I-C-MAUQ still retained the 21 items and 3 dimensions of the original MAUQ and the C-MAUQ: 8 items for usability and satisfaction, 6 items for system information arrangement, and 7 items for efficiency. Improper translations of all the 21 items in the C-MAUQ were modified through discussion among the whole research team.

Improvement of Cross-cultural Adaptation

The C-MAUQ has been adapted cross-culturally through experts’ comments and a prediction test [15]. Based on this adaptation and drawing on Conway et al.’s translatability assessment (TA) [16], this study further adapted the C-MAUQ by inviting a group of bilingual translators and health educators to assess the comprehensibility of the content as well as the cultural relevance and appropriateness of each item. Subsequently, the revised version was subjected to a trial survey online, in which 50 college students participated to identify problems that needed to be resolved.

Informants and Online Survey

Participants were students of the School of Foreign Studies, Nantong University, China. Impacted by varying degrees of psychological problems that became increasingly serious during the repeated COVID-19 attacks, these students urgently needed mHealth apps for self-diagnosis and general health information to relieve their psychologically strained minds. The questionnaire was administered using the online questionnaire survey platform named wenjuanxing [17] on February 18, 2022, and the survey lasted until no additional questionnaire was submitted online for 2 consecutive days (March 4, 2022). Over this period, the survey was announced to the entire student body of over 1000 at the School of Foreign Studies, Nantong University, through emails and WeChat groups. Meanwhile, the candidate informants were requested to use the Left-handed Doctor app for 2 days to become familiar with it before answering the questionnaire. The majority of participants in this study were female, which is characteristic of all schools of foreign studies in China.

Data Collection

The survey was conducted through wenjuanxing [17], the most popular online questionnaire platform in China. Two categories of data were collected via online questionnaires: the demographic information of the participants and their ratings on the 21 items concerning the usability of Left-handed Doctor. The demographic data included the informants’ age, gender, grade, and channel to obtain health information. The usability test elicited data concerning the informants’ ratings of the 21 items based on a 7-point Likert scoring system from 1 to 7 points (representing “strongly agree,” “agree,” “somewhat agree,” “neither agree nor disagree,” “somewhat disagree,” “disagree,” and “strongly disagree,” respectively).

Data Analysis

Quantitative analyses were conducted using SPSS version 22.0 (IBM, Inc.). First, demographic data were presented in a table and briefly described as the background information of the analysis. Subsequently, item analysis, weight analysis, and Pearson correlation analysis were conducted, followed by the reliability, validity, test-retest reliability, and split-half reliability tests. Finally, the range, mean values, and SD of the collected usability data were calculated and described for each of the 21 items.

Ethics Considerations

This study was approved and supported by the Student Affairs Office and the Humanities and Social Sciences Office of Nantong University, which is authorized to provide such approval before collecting data from students.

Results

Improvement of the C-MAUQ

Both translators (YS and MJ) found items 1, 2, 5, 9, 11-14, 17-21 problematic after checking the C-MAUQ against the original English version independently. They modified these items independently, and then, through discussion, agreed on the corresponding revisions and the classification of translation problems, which were subjected to further amendments before a final consensus among the study researchers. The translation problems in the C-MAUQ were related to (1) redundancy (items 1, 2, and 18); (2) incompleteness (item 12); (3) misuse of parts of speech (items 5, 9, and 17); (4) choice of inappropriate words (items 5, 9, 14, and 18-21); and (5) incomprehensibility (items 9, 11, and 13).

Further Cross-cultural Adaptation

The I-C-MAUQ was further adapted cross-culturally through a panel meeting attended by a group of bilingual translators and health educators. This meeting identified and agreed on a common problem concerning inappropriate cultural adaptation of items 18-21. In English-speaking countries, a patient always visits the same doctor and addresses the doctor as “my health care provider.” By contrast, in China, a patient usually sees different doctors when becoming ill and thus never uses “my” when referring to his/her “health care provider.” Therefore, “my” was crossed out from these 4 items. No other problems were detected during the panel meeting. After the panel meeting, the comprehensibility, readability, idiomaticity, and cultural adaptability of the questionnaire content were further improved. Subsequently, the I-C-MAUQ version was validated in an online trial survey completed by 50 informants. The trial survey turned out to be successful (Cronbach α=.992), and so the I-C-MAUQ did not require further improvement. The I-C-MAUQ, together with the C-MAUQ and the MAUQ, is provided in Multimedia Appendix 1.

Informant Demographics

Multimedia Appendix 2 shows the informants’ demographic information. A total of 322 responses were collected online, including 292 (90.7%) from female respondents. This can be explained by the fact that over 90% of students studying in the...
School of Foreign Studies, Nantong University, are females. The age of the participants ranged from 18 to 33 years (mean 21.68, SD 2.30 years). The overwhelming majority (n=316, 98.1%) were aged between 18 and 26 years. The informants included freshman (n=64, 19.9%), sophomore (n=29, 9.0%), junior (n=88, 27.3%), senior (n=48, 14.9%), first-year postgraduate candidates (n=46, 14.3%), and second-year postgraduate candidates (n=47, 14.6%). The majority of the informants (n=306, 95.0%) obtained health care information by visiting a doctor; logging into the internet; and communicating with families, friends, and classmates. Only a minor percentage of participants (n=9, 2.8%) used mHealth apps to obtain health care information.

**Questionnaire Item Analysis**

The 21 items in the I-C-MAUQ were valid and appropriately designed (Table 1), as evidenced by the distinction between the high-score group (n=94) and the low-score group (n=149). Data below the 27% quantile belonged to the low-score group, and those above the 73% quantile belonged to the high-score group. There was a significant difference in each of the 21 items between the high-score group and the low-score group, with P value in each case being <.001 (ie, P<.01). This indicates that all 21 items could well be distinguished from one another and thus should all be retained in the final version of the questionnaire. Besides, all the 21 items were significant (Table 2), with critical values (CR) determined within the range of 14.751-19.449 and the P value (CR) calculated at <.001 (ie, P<.01). The correlation coefficient between the score of each item and the total score of the questionnaire was determined within the range of 0.861-0.938 (<.01). Thus, all the 21 items were retained. According to the Pearson correlation values (Table S1 of Multimedia Appendix 3), all the 21 items were significantly and positively correlated, with the correlation coefficients ranging from 0.688 to 0.921 and P<.01.

**Table 1. Item analysis.**

<table>
<thead>
<tr>
<th>Items</th>
<th>Group, mean (SD)</th>
<th>t (critical values)</th>
<th>P value&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low-score group (n=149)</td>
<td>High-score group (n=94)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.58 (0.57)</td>
<td>3.65 (1.08)</td>
<td>17.031</td>
</tr>
<tr>
<td>2</td>
<td>1.52 (0.51)</td>
<td>3.34 (1.12)</td>
<td>14.751</td>
</tr>
<tr>
<td>3</td>
<td>1.56 (0.55)</td>
<td>3.54 (1.02)</td>
<td>17.251</td>
</tr>
<tr>
<td>4</td>
<td>1.58 (0.54)</td>
<td>3.66 (1.08)</td>
<td>17.352</td>
</tr>
<tr>
<td>5</td>
<td>1.58 (0.58)</td>
<td>3.51 (1.08)</td>
<td>15.961</td>
</tr>
<tr>
<td>6</td>
<td>1.57 (0.56)</td>
<td>3.53 (1.07)</td>
<td>16.348</td>
</tr>
<tr>
<td>7</td>
<td>1.58 (0.55)</td>
<td>3.85 (1.05)</td>
<td>19.449</td>
</tr>
<tr>
<td>8</td>
<td>1.56 (0.52)</td>
<td>3.62 (1.06)</td>
<td>17.555</td>
</tr>
<tr>
<td>9</td>
<td>1.61 (0.61)</td>
<td>3.60 (1.17)</td>
<td>15.225</td>
</tr>
<tr>
<td>10</td>
<td>1.53 (0.51)</td>
<td>3.61 (1.03)</td>
<td>18.186</td>
</tr>
<tr>
<td>11</td>
<td>1.52 (0.51)</td>
<td>3.49 (1.05)</td>
<td>16.905</td>
</tr>
<tr>
<td>12</td>
<td>1.58 (0.57)</td>
<td>3.46 (0.99)</td>
<td>16.724</td>
</tr>
<tr>
<td>13</td>
<td>1.52 (0.54)</td>
<td>3.40 (1.04)</td>
<td>16.262</td>
</tr>
<tr>
<td>14</td>
<td>1.56 (0.52)</td>
<td>3.65 (1.04)</td>
<td>18.038</td>
</tr>
<tr>
<td>15</td>
<td>1.55 (0.53)</td>
<td>3.55 (1.06)</td>
<td>16.993</td>
</tr>
<tr>
<td>16</td>
<td>1.56 (0.56)</td>
<td>3.56 (1.11)</td>
<td>16.178</td>
</tr>
<tr>
<td>17</td>
<td>1.56 (0.52)</td>
<td>3.51 (1.09)</td>
<td>16.242</td>
</tr>
<tr>
<td>18</td>
<td>1.55 (0.53)</td>
<td>3.54 (1.04)</td>
<td>17.182</td>
</tr>
<tr>
<td>19</td>
<td>1.55 (0.55)</td>
<td>3.50 (1.03)</td>
<td>16.832</td>
</tr>
<tr>
<td>20</td>
<td>1.67 (0.67)</td>
<td>3.68 (0.98)</td>
<td>17.527</td>
</tr>
<tr>
<td>21</td>
<td>1.56 (0.52)</td>
<td>3.57 (1.08)</td>
<td>16.862</td>
</tr>
</tbody>
</table>

<sup>a</sup>Items 1-21 represent the 21 items in the questionnaire.

<sup>b</sup>All P values <.01.
Table 2. Correlation between the 21 items and the overall score of the questionnaire.

<table>
<thead>
<tr>
<th>Items</th>
<th>CR&lt;sup&gt;a&lt;/sup&gt;</th>
<th>P value (CR)</th>
<th>COSQ&lt;sup&gt;b&lt;/sup&gt;</th>
<th>P value&lt;sup&gt;c&lt;/sup&gt; (COSQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17.031</td>
<td>&lt;.001</td>
<td>0.874</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2</td>
<td>14.751</td>
<td>&lt;.001</td>
<td>0.885</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>3</td>
<td>17.251</td>
<td>&lt;.001</td>
<td>0.902</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>4</td>
<td>17.352</td>
<td>&lt;.001</td>
<td>0.907</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>5</td>
<td>15.961</td>
<td>&lt;.001</td>
<td>0.861</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>6</td>
<td>16.348</td>
<td>&lt;.001</td>
<td>0.883</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>7</td>
<td>19.449</td>
<td>&lt;.001</td>
<td>0.890</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>8</td>
<td>17.555</td>
<td>&lt;.001</td>
<td>0.921</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>9</td>
<td>15.225</td>
<td>&lt;.001</td>
<td>0.879</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>10</td>
<td>18.186</td>
<td>&lt;.001</td>
<td>0.925</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>11</td>
<td>16.905</td>
<td>&lt;.001</td>
<td>0.938</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>12</td>
<td>16.724</td>
<td>&lt;.001</td>
<td>0.923</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>13</td>
<td>16.262</td>
<td>&lt;.001</td>
<td>0.906</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>14</td>
<td>18.038</td>
<td>&lt;.001</td>
<td>0.923</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>15</td>
<td>16.993</td>
<td>&lt;.001</td>
<td>0.914</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>16</td>
<td>16.178</td>
<td>&lt;.001</td>
<td>0.879</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>17</td>
<td>16.242</td>
<td>&lt;.001</td>
<td>0.910</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>18</td>
<td>17.182</td>
<td>&lt;.001</td>
<td>0.912</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>19</td>
<td>16.832</td>
<td>&lt;.001</td>
<td>0.896</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>20</td>
<td>17.527</td>
<td>&lt;.001</td>
<td>0.869</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>21</td>
<td>16.862</td>
<td>&lt;.001</td>
<td>0.905</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>CR: critical value.

<sup>b</sup>COSQ: correlation with the overall score of the questionnaire.

<sup>c</sup>All P values <.01.

Weight of the 21 Items in the Questionnaire

Through the analytic hierarchy process, the weight of each of the 21 items in the questionnaire was determined. Based on the judgment matrix of the 21 items (Table S2 of Multimedia Appendix 3), the eigenvector and weight of each item were determined (Table 3). Drawing on the eigenvectors, the maximum eigenvalue (21.000) was worked out. According to the maximum eigenvalue, the CI (<0.001) was computed. According to Table 4, the random index (RI) of the judgment matrix was 1.6358. From the CI (<0.001) and the RI (1.6358), CR (<0.001) was finally calculated (Table 5). This CR value (<0.1) indicated that the judgment matrix passed the consistency test. Therefore, the weights of the 21 items in Table 3 were valid. These weight values meant that the 21 items were almost equally important in the questionnaire.
### Table 3. Analytic hierarchy process analysis of the 21 items in the questionnaire.

<table>
<thead>
<tr>
<th>Items</th>
<th>Eigenvectors</th>
<th>Weight (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.018</td>
<td>4.846</td>
</tr>
<tr>
<td>2</td>
<td>0.954</td>
<td>4.541</td>
</tr>
<tr>
<td>3</td>
<td>0.990</td>
<td>4.712</td>
</tr>
<tr>
<td>4</td>
<td>1.010</td>
<td>4.808</td>
</tr>
<tr>
<td>5</td>
<td>1.020</td>
<td>4.858</td>
</tr>
<tr>
<td>6</td>
<td>1.007</td>
<td>4.795</td>
</tr>
<tr>
<td>7</td>
<td>1.071</td>
<td>5.099</td>
</tr>
<tr>
<td>8</td>
<td>0.990</td>
<td>4.712</td>
</tr>
<tr>
<td>9</td>
<td>1.015</td>
<td>4.833</td>
</tr>
<tr>
<td>10</td>
<td>0.992</td>
<td>4.725</td>
</tr>
<tr>
<td>11</td>
<td>0.956</td>
<td>4.554</td>
</tr>
<tr>
<td>12</td>
<td>0.980</td>
<td>4.668</td>
</tr>
<tr>
<td>13</td>
<td>0.956</td>
<td>4.554</td>
</tr>
<tr>
<td>14</td>
<td>1.015</td>
<td>4.833</td>
</tr>
<tr>
<td>15</td>
<td>0.994</td>
<td>4.731</td>
</tr>
<tr>
<td>16</td>
<td>0.983</td>
<td>4.681</td>
</tr>
<tr>
<td>17</td>
<td>0.996</td>
<td>4.744</td>
</tr>
<tr>
<td>18</td>
<td>1.004</td>
<td>4.782</td>
</tr>
<tr>
<td>19</td>
<td>0.995</td>
<td>4.738</td>
</tr>
<tr>
<td>20</td>
<td>1.056</td>
<td>5.029</td>
</tr>
<tr>
<td>21</td>
<td>0.999</td>
<td>4.757</td>
</tr>
</tbody>
</table>

*Maximum eigenvalue: 21.000; CI < 0.001.*

### Table 4. RI<sup>a</sup> table of the judgment matrix.

<table>
<thead>
<tr>
<th>Order</th>
<th>RI</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.52</td>
<td>0.89</td>
<td>1.12</td>
<td>1.26</td>
<td>1.36</td>
<td>1.41</td>
<td>1.46</td>
<td>1.49</td>
<td>1.52</td>
<td>1.54</td>
<td>1.56</td>
<td>1.58</td>
<td>1.59</td>
<td>1.5943</td>
<td></td>
</tr>
<tr>
<td>Order</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>RI</td>
<td>1.6064</td>
<td>1.6133</td>
<td>1.6207</td>
<td>1.6292</td>
<td>1.6358</td>
<td>1.6403</td>
<td>1.6462</td>
<td>1.6497</td>
<td>1.6556</td>
<td>1.6587</td>
<td>1.6631</td>
<td>1.6670</td>
<td>1.6693</td>
<td>1.6724</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>RI: random index.

### Table 5. Consistency test of the weight of the 21 items.

<table>
<thead>
<tr>
<th>Maximum eigenvalue</th>
<th>CI</th>
<th>RI&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Critical value</th>
<th>Result of test</th>
</tr>
</thead>
<tbody>
<tr>
<td>21.000</td>
<td>&lt;0.001</td>
<td>1.636</td>
<td>&lt;0.001</td>
<td>Pass</td>
</tr>
</tbody>
</table>

<sup>a</sup>RI: random index.

### Questionnaire Reliability and Validity

The statistics in Table 6 indicate the high reliability of the questionnaire. The corrected item-total correlation values of the 21 items all fell within 0.845-0.931, far exceeding 0.4. This meant that the 21 items were strongly correlated, and that they all had a high degree of reliability. Besides, the Cronbach α did not apparently increase when each of the 21 items was deleted, which implied that all items should be retained in the questionnaire. The overall Cronbach α (.988) for the 21 items was well above 0.9, indicating that the data collected for each item in the questionnaire were highly reliable. The values of test-retest reliability and split-half reliability were 0.918 and 0.828, respectively. Therefore, all the data were suitable for further analysis.
Table 6. Questionnaire reliability (and internal consistency).

<table>
<thead>
<tr>
<th>Items</th>
<th>Corrected item-total correlation</th>
<th>Cronbach α if item deleted&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.860</td>
<td>.988</td>
</tr>
<tr>
<td>2</td>
<td>0.873</td>
<td>.988</td>
</tr>
<tr>
<td>3</td>
<td>0.891</td>
<td>.988</td>
</tr>
<tr>
<td>4</td>
<td>0.897</td>
<td>.988</td>
</tr>
<tr>
<td>5</td>
<td>0.845</td>
<td>.988</td>
</tr>
<tr>
<td>6</td>
<td>0.870</td>
<td>.988</td>
</tr>
<tr>
<td>7</td>
<td>0.877</td>
<td>.988</td>
</tr>
<tr>
<td>8</td>
<td>0.912</td>
<td>.987</td>
</tr>
<tr>
<td>9</td>
<td>0.866</td>
<td>.988</td>
</tr>
<tr>
<td>10</td>
<td>0.917</td>
<td>.987</td>
</tr>
<tr>
<td>11</td>
<td>0.931</td>
<td>.987</td>
</tr>
<tr>
<td>12</td>
<td>0.915</td>
<td>.987</td>
</tr>
<tr>
<td>13</td>
<td>0.896</td>
<td>.988</td>
</tr>
<tr>
<td>14</td>
<td>0.914</td>
<td>.987</td>
</tr>
<tr>
<td>15</td>
<td>0.905</td>
<td>.987</td>
</tr>
<tr>
<td>16</td>
<td>0.866</td>
<td>.988</td>
</tr>
<tr>
<td>17</td>
<td>0.900</td>
<td>.987</td>
</tr>
<tr>
<td>18</td>
<td>0.902</td>
<td>.987</td>
</tr>
<tr>
<td>19</td>
<td>0.885</td>
<td>.988</td>
</tr>
<tr>
<td>20</td>
<td>0.855</td>
<td>.988</td>
</tr>
<tr>
<td>21</td>
<td>0.895</td>
<td>.988</td>
</tr>
</tbody>
</table>

<sup>a</sup>Cronbach α (standardized)=.988.

Table 7 reveals that the questionnaire is highly valid. The communalities for all 21 items ranged from 0.738 to 0.881, well above 0.4, indicating that the data can effectively be extracted from all these items. The Kaiser–Meyer–Olkin (KMO) value (0.973) was above 0.9, which showed that all the data concerning the 21 items could effectively be extracted. The percentage of variance (rotated) for factor 1 was 81.053%, considerably above 50%, meaning that all the data on all the items can validly be extracted.
Table 7. Questionnaire validity.

<table>
<thead>
<tr>
<th>Items</th>
<th>Factor loadings (factor 1)</th>
<th>Communalitiesa</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.873b</td>
<td>0.762</td>
</tr>
<tr>
<td>2</td>
<td>0.885b</td>
<td>0.784</td>
</tr>
<tr>
<td>3</td>
<td>0.902b</td>
<td>0.813</td>
</tr>
<tr>
<td>4</td>
<td>0.907b</td>
<td>0.822</td>
</tr>
<tr>
<td>5</td>
<td>0.859b</td>
<td>0.738</td>
</tr>
<tr>
<td>6</td>
<td>0.882b</td>
<td>0.778</td>
</tr>
<tr>
<td>7</td>
<td>0.889b</td>
<td>0.790</td>
</tr>
<tr>
<td>8</td>
<td>0.921b</td>
<td>0.848</td>
</tr>
<tr>
<td>9</td>
<td>0.878b</td>
<td>0.771</td>
</tr>
<tr>
<td>10</td>
<td>0.925b</td>
<td>0.856</td>
</tr>
<tr>
<td>11</td>
<td>0.939b</td>
<td>0.881</td>
</tr>
<tr>
<td>12</td>
<td>0.924b</td>
<td>0.854</td>
</tr>
<tr>
<td>13</td>
<td>0.907b</td>
<td>0.823</td>
</tr>
<tr>
<td>14</td>
<td>0.923b</td>
<td>0.852</td>
</tr>
<tr>
<td>15</td>
<td>0.915b</td>
<td>0.837</td>
</tr>
<tr>
<td>16</td>
<td>0.880b</td>
<td>0.774</td>
</tr>
<tr>
<td>17</td>
<td>0.911b</td>
<td>0.830</td>
</tr>
<tr>
<td>18</td>
<td>0.912b</td>
<td>0.832</td>
</tr>
<tr>
<td>19</td>
<td>0.896b</td>
<td>0.803</td>
</tr>
<tr>
<td>20</td>
<td>0.868b</td>
<td>0.754</td>
</tr>
<tr>
<td>21</td>
<td>0.905b</td>
<td>0.819</td>
</tr>
<tr>
<td>Eigenvalues (initial)</td>
<td>17.021</td>
<td>N/Ac</td>
</tr>
<tr>
<td>Variance (%) (initial)</td>
<td>81.053</td>
<td>N/Ac</td>
</tr>
<tr>
<td>Cumulative variance (%) (initial)</td>
<td>81.053</td>
<td>N/Ac</td>
</tr>
<tr>
<td>Eigenvalues (rotated)</td>
<td>17.021</td>
<td>N/Ac</td>
</tr>
<tr>
<td>Variance (%) (rotated)</td>
<td>81.053</td>
<td>N/Ac</td>
</tr>
<tr>
<td>Cumulative variance (%) (rotated)</td>
<td>81.053</td>
<td>N/Ac</td>
</tr>
<tr>
<td>Kaiser–Meyer–Olkin</td>
<td>0.973</td>
<td>N/Ac</td>
</tr>
<tr>
<td>Bartlett test of sphericity (chi-square); df</td>
<td>10873.765; 210</td>
<td>N/Ac</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>N/Ac</td>
</tr>
</tbody>
</table>

aThe communality is less than 0.4.
bThe absolute value of loading is greater than 0.4.
cN/A: not applicable.

Usability of the Left-handed Doctor App

Table 8 presents the results of the descriptive analysis of the usability of Left-handed Doctor. The range, mean (SD), and median scores were based on the rating of each item (1=strongly agree; 2=agree; 3=somewhat agree; 4=neither agree nor disagree; 5=somewhat disagree; 6=disagree; and 7=strongly disagree). The mean scores of the 21 items were between 2.224 and 2.497, indicating that the respondents were inclined to agree.
with the statements in all 21 items. In other words, they found the Left-handed Doctor app usable on the whole.

There were no significant differences ($P=.35$) in the mean scores concerning the 3 dimensions of usability and satisfaction (items 1-8), the arrangement of system information (items 9-14), and efficiency (items 15-21). This implied that the participants found the Left-handed Doctor app equally usable when it comes to the 3 dimensions.

Multimedia Appendix 4 shows the proportion of respondents falling into each of the 7 ratings of the 21 items. Over 60% (205/322, 63.7%; 223/322, 69.3%; 209/322, 64.9%; 206/322, 64.0%; 199/322, 61.8%; 206/322, 64.0%; 210/322, 65.2%; 203/322, 63.0%; 208/322, 64.6%; 219/322, 68.0%; 211/322, 65.5%; 218/322, 67.7%; 198/322, 61.5%; 208/322, 64.6%; 216/322, 67.1%; 207/322, 64.3%; 198/322, 61.5%; 203/322, 63.0%; 208/322, 64.6%, for items 1-6, 8-19, and 21, respectively) of informants strongly agreed or agreed with all items but items 7 (183/322, 56.8%) and 20 (187/322, 58.1%). Over 80% (267/322, 82.9%; 285/322, 88.5%; 277/322, 86.0%; 277/322, 86.0%; 259/322, 80.4%; 282/322, 87.6%; 270/322, 83.9%; 279/322, 86.6%; 287/322, 89.1%; 285/322, 88.5%; 288/322, 89.4%; 277/322, 86.0%; 280/322, 87.0%; 280/322, 87.0%; 276/322, 85.7%; 276/322, 85.7%; 258/322, 80.1%; 277/322, 86.0%, for items 1-21, respectively) of participants strongly agreed, agreed, or somewhat agreed with all the 21 items. This meant that the vast majority of the participating students showed a positive attitude toward the usability of the Left-handed Doctor app.

| Table 8. Descriptive analysis of the usability of the Left-handed Doctor app. |
|---------------------------------|----------------|-----------------|----------------|
| Item | Samples, n | Range | Mean (SD) | Median |
| 1 | 322 | 1.000-7.000 | 2.373 (1.180) | 2.000 |
| 2 | 322 | 1.000-7.000 | 2.224 (1.079) | 2.000 |
| 3 | 322 | 1.000-7.000 | 2.307 (1.125) | 2.000 |
| 4 | 322 | 1.000-7.000 | 2.354 (1.160) | 2.000 |
| 5 | 322 | 1.000-7.000 | 2.379 (1.176) | 2.000 |
| 6 | 322 | 1.000-7.000 | 2.348 (1.170) | 2.000 |
| 7 | 322 | 1.000-7.000 | 2.497 (1.246) | 2.000 |
| 8 | 322 | 1.000-7.000 | 2.307 (1.136) | 2.000 |
| 9 | 322 | 1.000-7.000 | 2.366 (1.182) | 2.000 |
| 10 | 322 | 1.000-7.000 | 2.314 (1.140) | 2.000 |
| 11 | 322 | 1.000-7.000 | 2.230 (1.101) | 2.000 |
| 12 | 322 | 1.000-7.000 | 2.286 (1.070) | 2.000 |
| 13 | 322 | 1.000-7.000 | 2.230 (1.089) | 2.000 |
| 14 | 322 | 1.000-7.000 | 2.366 (1.151) | 2.000 |
| 15 | 322 | 1.000-7.000 | 2.317 (1.132) | 2.000 |
| 16 | 322 | 1.000-7.000 | 2.292 (1.153) | 2.000 |
| 17 | 322 | 1.000-7.000 | 2.323 (1.117) | 2.000 |
| 18 | 322 | 1.000-7.000 | 2.342 (1.125) | 2.000 |
| 19 | 322 | 1.000-7.000 | 2.340 (1.199) | 2.000 |
| 20 | 322 | 1.000-7.000 | 2.463 (1.166) | 2.000 |
| 21 | 322 | 1.000-7.000 | 2.329 (1.137) | 2.000 |

Discussion

Principal Findings

Informed by Zhou et al [9] and Mustafa et al [13], the study improved the C-MAUQ translated, adapted, and validated in Zhao et al [14], and then used the I-C-MAUQ to test the usability of Left-handed Doctor, one of the most popular “reaching out to patients” interactive mHealth apps in China. The I-C-MAUQ had a better internal consistency (the correlation coefficient between the score of each item and the total score of the questionnaire ranging from 0.861 to 0.938; $P<.001$), reliability (Cronbach $\alpha=.988$), validity (load factor ranging from 0.859 to 0.939, percentage of cumulative variance [rotated]=81.053%, $KMO=0.973$), test-retest reliability (0.918), and split-half reliability (0.828) than the C-MAUQ [14]. Such better performance of the I-C-MAUQ resulted from 4 factors: (1) better comprehensibility, readability, and cultural adaptation of the I-C-MAUQ; (2) different categories of participants in terms of age, gender, education, profession, and sample size; (3) different functions of the tested interactive mHealth apps used by patients (with vs without the chatbot function); and (4) respondents’ experience with mHealth apps. Similarly, we found...
that the reliability of the I-C-MAUQ was relatively higher than those reported in Mustafa et al [13] (Cronbach α=.946; corrected item-total correlation values between –0.057 and 0.868) and Zhou et al [9] (Cronbach α=.914). We once again attributed the reliability difference to the aforesaid 4 factors, which will be discussed in the following sections.

**Cross-cultural Adaptation of the Translated Questionnaire**

It is imperative to adapt questionnaires cross-culturally, but there is a lack of evidence for the best approaches to cross-cultural adaptation (CCA) [18]. The most adopted methods for CCA are Brislin’s Translation Model [19], the use of panels or committees [20-26], and focus groups [27]. However, this study adopted another effective but a commonly neglected model: TA [16]. Drawing on the cross-cultural issues proposed in TA, we improved the C-MAUQ [15] by making further cultural and linguistic adaptations, solving the translation problems concerning redundancy, incompleteness, misuse of parts of speech, choice of inappropriate words, incomprehensibility, and relevance and appropriateness on the cultural, semantic, syntactic, and pragmatic facets. The newly adapted questionnaire was equivalent to the original questionnaire [18]. TA thus makes it possible to identify alternative versions for translation purposes, modify original versions to optimize subsequent translation efforts, and detect and discuss irrelevant or inappropriate items early [16]. Thus, TA needs to be adopted as an effective CCA method in prospective translation and adaptation of questionnaires.

**Participant Differences in Age, Gender, Education, Profession, and Sample Size**

Most (318/322, 98.8%) of the informants in this study were aged 18-28, compared with the majority (91.04%) of respondents aged 29-65 in Zhao et al [14], with just over half (52.3%) of the participants aged 18-28 and just below half (48.3%) aged 29-65 in Zhou et al [9], and with all (100%) those surveyed aged 22-25 in Mustafa et al [13]. We concluded that younger age potentially led to relatively positive ratings of questionnaire items and thus higher questionnaire reliability and internal consistency.

The proportions of male and female participants (30/322, 9.3% vs 291/322, 90.4%) were different from those (53.76% vs 46.24%) in Zhao et al [14], those (38.3% vs 61.7%) in Zhou et al [9], and those (8% vs 92%) in Mustafa et al [13]. Therefore, considerably higher percentages (292/322, 90.7%) of female respondents seemed to contribute to a higher degree of the questionnaire’s internal consistency and reliability. This result showed that females were more interested in participating in surveys on the usability of mHealth apps and that more female users of mobile apps were keen on using mHealth apps for health care. This has been also testified by Zhou et al [9].

All informants in this study and Mustafa et al [13] were college students at the undergraduate or graduate level, but those in Zhao et al [14] and Zhou et al [9] had different levels of education: 33.24% and 67.2% held an undergraduate or above in Zhao et al [14] and Zhou et al [9], respectively. The overall higher level of respondent education may explain the relatively higher degree of questionnaire’s internal consistency and reliability in our study and Mustafa et al [13], in comparison with that in Zhao et al [14] and Zhou et al [9]. However, the vast gap in participant education at or above the undergraduate level between Zhao et al [14] and Zhou et al [9] merely resulted in a considerably minor difference in questionnaire reliability (Cronbach α=.912 vs .914).

In terms of profession, being a student—100% (322/322) in this study and Mustafa et al [13], 31.4% in Zhou et al [9], and 1.56% in Zhao et al [14]—also likely impacted the questionnaire’s internal consistency and reliability, with the rate of students participating positively proportional to the degree of reliability and internal consistency.

These findings concerning age, gender, education, and profession contradicted the result in Zhou et al [9], which asserted that the demographic factors (eg, age, gender, education, occupation) failed to significantly impact the answers to the individual statements or the overall score on the MAUQ. The sample size was indeed not a contributing factor to the high internal consistency and reliability of the questionnaire. Zhao et al [14] recruited the largest number of participants (n=346) but reported the lowest internal consistency and reliability, whereas this study achieved the highest internal consistency and reliability of the questionnaire based on the data contributed by a similar number of informants (n=322), followed by a slightly lower internal consistency and reliability derived from the information provided by the smallest number of informants in Mustafa et al [13].

**Respondents’ Experience With mHealth Apps**

The informants in Zhou et al [9] used mobile apps for an average of 6.64 years; 86.42% of participants in Zhao et al [14] used mHealth apps more than 3 times during the month before the survey. Only 2.8% (9/322) of respondents in this study reported using mHealth apps for health care information, but they were requested to install the Left-handed Doctor app 2 weeks beforehand to become familiar with it. The informants in Mustafa et al [13] were also asked to do the same. Therefore, experience with mHealth apps did not seem to influence the users’ perceived usability, and thus the internal consistency and reliability of the questionnaire adopted remained unaffected.

**Interactive mHealth Apps for Patients Equipped With or Without the Chatbot Function**

This study tested the usability of the I-C-MAUQ on the Left-handed Doctor app, which is empowered with the chatbot function. By contrast, Zhao et al [14] adopted the Good Doctor app, which was not equipped with the chatbot function. This difference in apps may somewhat explain the notable discrepancy in the questionnaire’s internal consistency and reliability between this study (Cronbach α=.988) and that by Zhao et al [14] (Cronbach α=.912). The mHealth apps used in Krebs and Duncan [7] and Mustafa et al [13] did not have the chatbot function. Thus, further research needs to be conducted to pinpoint the impact of this function on the usability of mHealth apps.
Implications
It is worth adapting established and appropriate questionnaires with recorded validity because designing a new one is effort-, time-, and cost-consuming [12]. Proper translation and adaptation and TA [16,28] are essential to ensure equivalence between the original questionnaire and the translated version. Cultural and linguistic sensitivity is a prerequisite for ironing out the translation problems resulting from cultural and linguistic differences and making the translated questionnaire culturally relevant and appropriate. Therefore, qualified translators highly proficient in the source and target languages and health educators or practitioners need to make joint efforts to complete this challenging task.

Validation is crucial for ensuring the equivalence between the original version and the translated one. Content validity index has been used to quantify the questionnaire validity in some studies [9,13,15,29,30]. It has been widely used because of its simple measurement, accessibility, power to provide details for each item, and indication of item modification or deletion [30].

Limitations
This study has several limitations. First, the convenient sampling of college students from a single university made it challenging to generalize the findings to the whole population in China. The recruitment of only healthy students also made the generalization of the results less convincing. Finally, the sample size was not sufficiently large to guarantee the generalization of findings.

Conclusions
The I-C-MAUQ is highly reliable and valid for the Left-handed Doctor app, and thus suitable for testing the usability of interactive mHealth apps used by patients in China. This finding is in line with the study by Marzuki et al [12], further confirming the cross-cultural validity, reliability, and adaptability of the MAUQ. We identified certain factors that influence the perceived usability of mHealth apps, including users' age, gender, education, profession, and possibly previous experience with mHealth apps as well as the chatbot function of such apps. Most notably, we found a wider acceptance of this new technology among young Chinese female college students who were more engaged in the interaction with health care chatbots. The age-, gender- and profession-induced preference for new digital health interventions in China aligns with the findings from other similar studies in the United States [9] and Malaysia [13]. This preference identifies areas for further research on the social, cultural, and gender adaptation of health technologies.

Conflicts of Interest
None declared.

Multimedia Appendix 1
I-C-MAUQ, together with the C-MAUQ and the MAUQ. See also [31].
[DOCX File , 62 KB - formative_v6i7e37933_app1.docx ]

Multimedia Appendix 2
Informants' demographics.
[DOCX File , 16 KB - formative_v6i7e37933_app2.docx ]

Multimedia Appendix 3
Additional tables.
[DOCX File , 32 KB - formative_v6i7e37933_app3.docx ]

Multimedia Appendix 4
Frequency analysis of the usability of the Left-handed Doctor app.
[DOCX File , 29 KB - formative_v6i7e37933_app4.docx ]

References


17. wenjuanxing. wxj. URL: https://www.wxj.cn/ [accessed 2022-06-07]


Abbreviations

CCA: cross-cultural adaptation
CITC: the corrected item-total correlation
C-MAUQ: the Chinese version of the MAUQ
CR: critical value
I-C-MAUQ: the improved C-MAUQ
MAUQ: mHealth App Usability Questionnaire
mHealth: mobile health
PSSUQ: Post-Study System Usability Questionnaire
RI: random index
SUS: System Usability Scale
TA: translatability assessment

©Yi Shan, Meng Ji, Wenxiu Xie, Rongying Li, Xiaobo Qian, Xiaomin Zhang, Tianyong Hao. Originally published in JMIR Formative Research (https://formative.jmir.org), 06.07.2022. 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.
Perspectives of Patients and Therapists on Social Media and Digital Data Use in Mental Health Therapy: Thematic Analysis

Lauren Southwick, MPH; Rebecca Suh, BA; Elissa Kranzler, PhD; Megan Bradley, BA; Raina M Merchant, MD

University of Pennsylvania, Philadelphia, PA, United States
Fors Marsh Group, Washington DC, DC, United States
Department of Emergency Medicine, Perelman School of Medicine, Philadelphia, PA, United States

*these authors contributed equally

Corresponding Author: Lauren Southwick, MPH
University of Pennsylvania
3600 Civic Center Blvd
Philadelphia, PA, 19014
United States
Phone: 1 9145826995
Email: lauren.southwick@pennmedicine.upenn.edu

Abstract

Background: Incorporating insights from social media into the patient-provider encounter is increasingly being explored in health care settings. Less is known about the utility of these data in mental health therapy.

Objective: This study aims to prospectively investigate and characterize how social media and digital data are used in mental health therapy from both the patient and mental health therapist perspective.

Methods: Patients enrolled in mental health therapy and mental health therapists were interviewed using a semistructured interview guide. All interviews were transcribed and coded using a deductive framework analysis. Themes and subthemes were identified. Participants completed a sociodemographic survey, while mental health therapists also completed a behavioral norms and elicitation survey.

Results: Seventeen participants, that is, 8 (48%) mental health therapists and 9 (52%) patients were interviewed. Overall, participants identified 4 themes and 9 subthemes. Themes were current data collection practices, social media and digital data in therapy, advantages of social media and digital data in therapy, and disadvantages of social media and digital data in therapy. Most subthemes were related to the advantages and disadvantages of incorporating digital data in mental health therapy. Advantage subthemes included convenience, objective, builds rapport, and user-friendliness while disadvantage subthemes were nonreflective, ethically ambiguous, and nongeneralizable. The mental health therapists’ behavioral norms and elicitation survey found that injunctive and descriptive normative beliefs mapped onto 2 advantage subthemes: convenience and objectivity.

Conclusions: This qualitative pilot study established the advantages and disadvantages of social media and digital data use in mental health therapy. Patients and therapists highlighted similar concerns and uses. This study indicated that overall, both patients and therapists are interested in and are comfortable to use and discuss social media and digital data in mental health therapy.

Keywords: social media; digital health; digital data; mental health therapy; mobile phone

Introduction

Background

Adults in the United States are frequent users of social media platforms such as Facebook, Instagram, and Twitter [1,2]. Such platforms have a profound impact on everyday life and provide new opportunities for understanding behavioral, social, and environmental determinants of mental health and well-being [3-5]. Social media data (eg, Facebook wall posts) and digital data (eg, search engine use, step data, smartphone metadata) are increasingly used to support patient mental health care [6]. Previous mental health research has demonstrated that social
media data can reveal and predict risk for mental health conditions such as depression, loneliness, suicide ideation, posttraumatic stress, schizophrenia, and bipolar disorder [6-12]. Furthermore, prior research has demonstrated that data from these digital platforms can provide critical information not readily attained through in-person or remote health care encounters to help therapists identify, address, or discuss mental health concerns [3,13,14].

In response to these findings, several researchers are increasingly capturing how and how often therapists incorporate social media data into mental health therapy and treatment. For example, Fisher and Appelbaum [15] report that some mental health clinicians incorporate parts of their patients’ Facebook feeds in their care delivery, whereas Hobbs and colleagues [13] found that nearly two-thirds of outpatient psychotherapists report viewing at least one patient’s social or electronic media (such as email messages, SMS text messaging, and other messaging apps) as part of psychotherapy. Hobbs et al [13] also report that the psychotherapists who access their patients’ electronic or social media data indicated that it improved their ability to provide effective treatment. These findings underscore the utility of social media and digital data in mental health therapy. These examples reflect the growing body of research on the mental health care therapist experience using social media data in mental health therapy; however, research on the patient perspective remains limited. Accordingly, research exploring both patient and therapist perspectives on the use of social media data in mental health therapy is warranted.

**Objective**

This qualitative study aimed to provide new knowledge on mental health patient and therapist perspectives regarding the use of social media data in mental health therapy. The aims of this study were to (1) explore patients’ and therapists’ current use(s) of social media data in mental health therapy and (2) identify the advantages and disadvantages of sharing social media data in mental health therapy.

**Methods**

**Recruitment**

Individuals in mental health therapy (referred to as patients) were recruited through a clinical research registry at a large academic institution (a private university and medical center in the northeastern region of the United States) from March to May 2018. Thirteen applicants expressed interest from the registry, of which 9 were available for an in-person interview. Patient inclusion criteria included that they attended mental health therapy for anxiety or depression and were aged 18 years and older. Patients completed informed consent and received a US $20 gift card as compensation for their study participation. Mental health therapists (referred to as therapists) were recruited at the same large academic institution through behavioral health research consortiums and word-of-mouth sampling (approximately 50 received an email invitation) from October to December 2018. Twelve therapists expressed interest and 8 were available for an in-person interview. Therapist inclusion criteria included that they provided behavioral health care and worked with clients aged 18 years and older with depression or anxiety disorder(s). The therapists completed informed consent and received a US $100 gift card as compensation for study participation. All participants were willing to complete an in-person interview at a large academic hospital, which was held in a secure private room and lasted for 50-70 minutes.

**Qualitative Interview**

Two semistructured interview guides were constructed based on a review of the published literature on social media data in therapy and co-design techniques [15,16]. The interview guide was structured as follows: introduction, discussion of interview expectations, current use of data in therapy, experience with social media data (eg, Facebook, Instagram, Twitter) in therapy, and advantages and disadvantages of incorporating social media data in therapy. Interview questions were tailored to separately address the patient and therapist experiences. All interviews were conducted by 2 team members. One individual (a cisgender female researcher with >10 years of experience in qualitative interviewing) acted as the lead facilitator, while the other interviewer (a cisgender female service designer with 5 years of experience) took notes and contributed to probing and follow-up questions. Prompts were included, where appropriate, to elicit participant elaboration about each topic. The full interview guide is available from the authors upon request.

**Survey**

Patients completed a brief in-person survey at the end of the interview. The survey assessed sociodemographic characteristics such as gender, race/ethnicity, highest level of education, and the Social Media Use Questionnaire (SMUQ) [17]. The SMUQ assesses problematic use of social media and comprises 9 items (eg, “I feel anxious when I am not able to check my social network account”), with response options on a 5-point Likert scale from “Never” to “Always.” All items were averaged into a scale for which higher scores corresponded with excessive social media use.

At 24 hours after the interview, therapists received a “thank you” email and a link to a web-based survey, programmed in Qualtrics software. The survey was intended to elicit participants’ beliefs pertinent to using data about patients’ social media use in therapy sessions, given little existing research on social media in mental health therapy. Survey questions were drawn from established procedures for elicitation studies [18]. These open-ended questions elicited beliefs about using social media data in therapy sessions in the next month. Types of beliefs elicited included behavioral beliefs (ie, potential benefits or drawbacks of using social media data in therapy sessions), injunctive normative beliefs (ie, individuals or groups of people who may or may not approve of using social media data in therapy sessions), descriptive normative beliefs (ie, individuals or groups of people who may or may not use social media data in therapy sessions), and control beliefs (ie, circumstances that may help or hinder the use of social media data in therapy sessions). The survey also assessed the sociodemographic characteristics of the participants, including gender, race, ethnicity, highest level of education, and work environment characteristics (eg, type of practice, caseload).
Ethics Approval

This study protocol was reviewed and approved by the University of Pennsylvania Institutional Review Board (protocol 831246).

Analysis

Qualitative Interviews

Therapist interviews were audio recorded and transcribed verbatim. Patient interviews were not recorded; a research assistant observed the interviews and took detailed notes and collected quotes. After multiple readings, the transcripts and interview notes were then coded by authors LS and RS. The coders created a codebook using the interview guide as themes, whereas subthemes emerged from the transcripts. The transcripts and interview notes were then analyzed using a deductive framework analysis [19] and coded according to the 6 stages of framework analysis: familiarization, identifying a thematic framework, indexing, charting, mapping, and interpretation. Identified subthemes were deduced from the coded passages and placed into separate coded charts. The authors then independently reviewed the charts for consistency and agreement. The coders met regularly to resolve disagreements for any theme or subtheme by consensus following discussion. The codebook is available from the authors upon request.

Survey Findings

All quantitative analyses (ie, descriptive statistics) were conducted in a Microsoft application (Microsoft Excel, version 16.58).

Results

Study Sample

The sociodemographic characteristics of the patients and therapists are provided in Table 1. Patients were mostly female (6/9, 66%) and non-Hispanic White (6/9, 66%). Patients were aged 22-60 years; half reported having completed college while the other half completed graduate school. According to the SMUQ, all patients reported using social media platforms (eg, Facebook, Instagram) more than 3 times a day and exhibited nonproblematic social media use (mean score 22, range 9-32). All therapists were females and non-Hispanic White, with a mean age of 37 years. Most reported that they work in a hospital/medical setting and practice cognitive behavioral therapy (CBT). However, therapists were at various stages in their careers, ranging from graduate student to psychologist or associate professor (Table 1).

<table>
<thead>
<tr>
<th>Sociodemographic variables</th>
<th>Patients (n=9)</th>
<th>Range</th>
<th>Therapists (n=8)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>34</td>
<td>22-60</td>
<td>37</td>
<td>29-43</td>
</tr>
<tr>
<td>Female</td>
<td>6</td>
<td>N/A</td>
<td>8</td>
<td>N/A</td>
</tr>
<tr>
<td>Ethnicity/race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>6</td>
<td>N/A</td>
<td>8</td>
<td>N/A</td>
</tr>
<tr>
<td>Black</td>
<td>1</td>
<td>N/A</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>Asian</td>
<td>1</td>
<td>N/A</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>Non-Hispanic, Latinx</td>
<td>1</td>
<td>N/A</td>
<td>8</td>
<td>N/A</td>
</tr>
<tr>
<td>Highest level of education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College graduate</td>
<td>5</td>
<td>N/A</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>More than college graduate (eg, master’s, doctoral degree)</td>
<td>4</td>
<td>N/A</td>
<td>8</td>
<td>N/A</td>
</tr>
<tr>
<td>Social media use questionnaire</td>
<td>22</td>
<td>9-32</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Cognitive behavioral therapy orientation</td>
<td>N/A</td>
<td>N/A</td>
<td>8</td>
<td>N/A</td>
</tr>
<tr>
<td>Clinical experience (years)</td>
<td>N/A</td>
<td>N/A</td>
<td>5.3</td>
<td>2-9</td>
</tr>
<tr>
<td>Clinical setting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community outpatient</td>
<td>N/A</td>
<td>N/A</td>
<td>1</td>
<td>N/A</td>
</tr>
<tr>
<td>Hospital</td>
<td>N/A</td>
<td>N/A</td>
<td>4</td>
<td>N/A</td>
</tr>
<tr>
<td>Private practice</td>
<td>N/A</td>
<td>N/A</td>
<td>3</td>
<td>N/A</td>
</tr>
<tr>
<td>Caseload (clients)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>5-40</td>
</tr>
<tr>
<td>Prep time prior to session (minutes)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>3-30</td>
</tr>
</tbody>
</table>

aN/A: not applicable.
Thematic Areas

The interviews captured patients’ and therapists’ social media use in mental health therapy. The interviews identified the advantages and disadvantages of sharing social media data in mental health therapy and highlighted the contextual and logistical considerations to incorporating these new data. The interviews were structured on the following themes: (1) current data use in therapy, (2) experience with social media in therapy, (3) advantages of social media in therapy, and (4) disadvantages of social media in therapy (Table 2).
Table 2. Emerging interview subthemes with illustrative quote(s).

<table>
<thead>
<tr>
<th>Themes, subthemes</th>
<th>Illustrative quote(s)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current data use in therapy</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Applications</td>
<td>…I think it is a great idea because if you get the notification from the app and it is kind of fun, it's visually enticing, and people are more likely to do it. And then some of them have the built-in mindfulness tracking.</td>
<td></td>
</tr>
<tr>
<td>Surveys</td>
<td>…Implement surveys during treatment to see change over time in symptoms or quality for life.</td>
<td>…Surveys are infrequent, I attend a self-determined care model.</td>
</tr>
<tr>
<td><strong>Experience with social media in therapy</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient-initiated</td>
<td>…The patient brought up Snapchat and showed conversation around the video…It was sharing a sexting video with someone they were interested in and were having anxious thoughts about having done that and the individual saved/downloaded the video.</td>
<td>…I am comfortable sharing social media because it's already out there, and everyone puts their business out on social media.</td>
</tr>
<tr>
<td>Provider-initiated</td>
<td>…There’s a lot of videos on YouTube to use for exposures. So…if I have somebody that comes in with like a serious vomit fear or emetophobia or fear of like, could be anything like scary movies or clowns. We just go on YouTube and look stuff up.</td>
<td>…I trust my instincts [on what to share].</td>
</tr>
<tr>
<td><strong>Advantages of incorporating social media in therapy</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Convenience</td>
<td>…I would also love for there to be more technology where the person doesn’t have to enter anything themselves. That’s why the Fitbit is nice or this sleep app where you basically put under your pillow (and) it tracks your sleep… because requiring them to do any work when they’re already depressed and anxious [and it can be] a huge burden.</td>
<td>…It would be nice to look back [at previous posts].</td>
</tr>
<tr>
<td>Objective</td>
<td>…I think sleep data could be really useful…Like I’ve had a lot of patients who tell me that they sleep for like 3 hours and I’m like that’s not possible every single night. I think this could be useful.</td>
<td>…It would help flag my memory… keep me on track.</td>
</tr>
<tr>
<td>Builds rapport</td>
<td>…I don’t think it derailed anything,… mostly it’s very relevant and helpful. They’re willing to share, and I think it builds rapport. Often times it helps me to like really understand what they’re talking about.</td>
<td>…We talk about [text messages] as a “how are you?”</td>
</tr>
<tr>
<td>User-friendly</td>
<td>…It’s nice that it graphs it for you…I would use it in the beginning of session when I’m asking how they are feeling right now and how they feel how the last month has been to sort of see if their self-report in the moment lines up.</td>
<td>…If therapy is holistic having this data might not be bad.</td>
</tr>
<tr>
<td><strong>Disadvantages of incorporating social media in therapy</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonreflective</td>
<td>…I think that people sometimes present very differently in therapy than they might on social media. If I’m thinking of somebody who is super depressed or maybe [have] chronic mental health issues, I think that that person probably doesn’t post a lot on social media about their depression.</td>
<td>…I trust my instincts on what to share.</td>
</tr>
<tr>
<td>Ethically ambiguous</td>
<td>…For them to know that I have access to everything they post could make them feel pretty watched. They don’t have any privacy because anything they do online the therapist will see it and will pull it up on their portal and judge them for it.</td>
<td>…I want to disclose what I need to [in each session]…maybe it’s my eating disorder [that] has been more on my mind, not my substance abuse.</td>
</tr>
<tr>
<td>Nongeneralizable</td>
<td>…I think there are different social media personalities. There are some people who are very explicitly like “I don’t talk about politics on Facebook.” And other people who use it as their diary where they are posting every thought. And then people who don’t post very much because they’re more private.</td>
<td>…My phone is so personal and feels like an invasion of privacy. It feels too “big brother.”</td>
</tr>
<tr>
<td></td>
<td>…I would like to [annotate them] (ie, text message or social media post) because things are left out.</td>
<td>…I am not confident in accuracy of an algorithm.</td>
</tr>
<tr>
<td></td>
<td>…I don’t want to be judged by my digital data.</td>
<td></td>
</tr>
</tbody>
</table>
Current Data Use in Therapy

When asked about current data collection practices, both patients and therapists noted using mental health and well-being apps in their sessions. Apps were mainly used during their practice to augment mental health services. For example, a patient reported using an app designed for patients with bipolar disorders I and II, depression, posttraumatic stress disorder, and anxiety disorder to keep track of symptoms and triggers (Table 2). However, several therapists expressed concern with selecting the “right” app, whereas others noted security and confidentiality issues for their patients. One therapist said, “[My patient] didn’t want to put the app in her phone because her friends look at her phone and she thought that they would ask about it” (Participant #1, female, aged 42 years). In addition to using apps, therapists reported that they routinely collect patient-reported outcome data. As a common practice of CBT, validated surveys such as Beck’s Depression Inventory, Patient Health Questionnaire 9-item, or Generalized Anxiety Disorder 7-item scales are collected at every therapy visit to track progress. Patients in this sample did not report completing validated surveys at their therapy sessions.

Experience Discussing Social Media in Therapy

The patients interviewed reported sharing and discussing social media data with their therapist. Most patients reported sharing or discussing Instagram posts (n=10), Facebook posts or statuses (n=8), tweets (ie, Twitter posts) (n=3), and YouTube videos (n=3). One patient noted, “I show my therapist my Facebook, it’s evidence of my life; yes, I have friends, I exist” (Participant #3, genderqueer, aged 24 years).

Approximately half of the patients reported that they have directly shown or summarized social media posts to their therapist in the last month. Therapists noted similar interactions; one elaborated as follows, “So for [instance] a patient that is dating and wants to show me some of the people that she’s talking about… she’ll show me on social media. Or they want to show me what they post, a story that’s going, something that they are involved with, like a family member or friends” (Participant #2, female, aged 35 years). Once social media data are introduced in therapy sessions, both patients and therapists detailed how social media platforms such as Facebook or YouTube can also be used as homework or in exposure therapy. For example, 1 patient explained that their therapist would have them write on a friend’s Facebook wall to address their social anxiety and phobia.

Advantages of Incorporating Social Media in Therapy

Both patients and therapists highlighted how social media data can be convenient, objective, user-friendly, and allow them to build rapport when used in therapy (Table 2). Both patients and therapists focused on how data are shared, may it be automatic or manual and how the algorithm could select posts as an advantage. Patients indicated a general acceptance of an algorithm selecting their social media posts but emphasized a desire to annotate or provide context to the post. One patient said, “I need to be there with [my therapist] to review” (Participant #5, male, aged 34 years). Therapists highlighted a desire for objective metrics derived from social media posts. They commonly referenced social media’s metadata such as time of post and language used. For example, therapists noted that patients recounting of events may be influenced by recall bias when individuals have a partial account of prior events [20], whereas seeing metadata provides an objective source of information. Both therapists and patients highlighted their comfort with and interest in digital data such as steps walked via a smartphone’s built-in pedometer or screen time metrics.

Social media data are seen to aid discussions and accelerate a patient’s account of events. Both participants and therapists highlighted how they would like to see trends over time. A therapist said, “I’m asking how they are feeling right now and how they feel how the last month has been to sort of see if their self-report in the moment lines up” (Participant #2, female, aged 35 years). Therapists also noted instances when their patient would show pictures of friends or family from a social media platform to add a “face to the name,” implying that the use of such data builds rapport, which in turn could enhance the patient-provider communication.

Disadvantages of Incorporating Social Media in Therapy

Our results underscored perceived disadvantages of using social media data in therapy. Specifically, it can be nonreflective, nongeneralizable, and its use could be ethically ambiguous. Patients reported a sense of fear and uneasiness “always” sharing social media data with their therapist. They noted concerns about being “watched” and saw it as “a little creepy.” Few therapists expressed concern that it could also elicit a Hawthorne effect, altering one’s behavior due to the awareness of being observed [21]. As 1 therapist noted, “And again, me seeing all of [the] posts patients put up, even if they are agreeing to that…Patients may change [the way] they engage with that social media platform” (Participant #6, female, aged 43 years).

Therapists expressed concern that sharing social media data is nonreflective and would not provide accurate depictions of the patient’s true thoughts and emotions. They highlighted how social media posts are often public-facing accounts of people or events and may not be genuinely authentic. One patient said that “things are left” out of posts, whereas therapists highlighted how their patients may have a social media personality.

Several participants raised important questions regarding the security of the data collected, with 1 patient saying, “I want to protect [my] autonomy” (Participant #7, female, aged 29 years). Similarly, a patient highlighted that they trust their intuition on what social media posts to share with their therapist (Participant #7, female, aged 29 years). Half of the therapists expressed concern regarding the social media platforms’ security policies. Several therapists also indicated that consistently including social media data in their therapy sessions could negatively impact their workflow.

Furthermore, utility of social media data in therapy may hinge on a patient’s age or comfort with technology. Both therapists and patients agreed that including these data would be the most beneficial for younger or more technologically inclined patients. One participant noted that social media in their therapy session may not be “beneficial for me [and my treatment goals]
but for a younger generation because they post so much” (Participant #6, female, aged 60 years).

**Additional Therapist Beliefs About Social Media Use in Therapy**

After the interviews concluded, therapists were asked a series of open-ended survey questions to assess their normative and control beliefs relevant to using patients’ social media data in therapy sessions. See Table 3 for the elicitation questions asked and the illustrative quotes. Therapists were asked to indicate what types of individuals or groups of people would be more or less likely to use social media in therapy (descriptive normative beliefs). Responses included being a younger therapist or patient and a digital native and someone who grew up with technology who might be more technologically inclined. Therapists indicated that their colleagues who would be open to this type of data exchange must be made aware of the social media platform themselves and be oriented toward CBT or other measurement-based care orientations. Therapists reported that individuals or groups who would be less likely to use social media in therapy included those who are older, have limited experience with social media platforms, and do not use measurement-based care. When asked to indicate what circumstances would make it difficult or easy to use social media in therapy (control beliefs), responses underscored the importance of convenience of use for both the therapist and the patient and ease of understanding in the context of therapy.
Table 3. Examples of open-ended responses and the corresponding theme categorization for each type of belief elicitation question.

<table>
<thead>
<tr>
<th>Belief category, elicitation questions, illustrative open-ended responses</th>
<th>Subtheme(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Normative beliefs (injunctive)</strong></td>
<td></td>
</tr>
<tr>
<td>Generally, what types of individuals or groups would approve or think you should use data about clients’ social media use in therapy sessions in the next month? Please list general groups or personas; do not include specific names.</td>
<td></td>
</tr>
<tr>
<td>Younger clinicians, clinicians who work with young adults, clinicians who use a measurement-based care framework (ie, track their clients’ progress using measures)</td>
<td>Convenience</td>
</tr>
<tr>
<td>Younger therapists, data-driven/number-oriented people</td>
<td>Objective</td>
</tr>
<tr>
<td>Generally, what types of individuals or groups would disapprove or think you should not use data about clients’ social media use in therapy sessions in the next month? Please list general groups or personas; do not include specific names.</td>
<td></td>
</tr>
<tr>
<td>Potentially psychodynamic practitioners, individuals with strong privacy concerns</td>
<td>Ethically ambiguous</td>
</tr>
<tr>
<td>Individuals who do not use or have social media, clients who may be mistrusting or not have a strong therapeutic rapport with their therapist</td>
<td>Builds rapport</td>
</tr>
<tr>
<td>Ethically ambiguous</td>
<td></td>
</tr>
<tr>
<td><strong>Normative beliefs (descriptive)</strong></td>
<td></td>
</tr>
<tr>
<td>Generally, what types of individuals or groups are most likely to use data about clients’ social media use in therapy sessions in the next month? Please list general groups or personas; do not include specific names.</td>
<td></td>
</tr>
<tr>
<td>Younger, more number-oriented practitioners</td>
<td>Convenience</td>
</tr>
<tr>
<td>Clinicians who are familiar with and comfortable using social media, clinicians who treat young adults, clinicians who incorporate technology into their treatments (eg, give measures on a computer or iPad, email, or text their clients)</td>
<td>Objective</td>
</tr>
<tr>
<td>Provider-initiated</td>
<td></td>
</tr>
<tr>
<td>Generally, what types of individuals or groups are least likely use data about clients’ social media use in therapy sessions in the next month? Please list general groups or personas; do not include specific names.</td>
<td></td>
</tr>
<tr>
<td>Older clinicians: clinicians in an environment in which it is inconvenient to do so</td>
<td>Convenience</td>
</tr>
<tr>
<td>Clinicians who do not use social media themselves and may have limited knowledge about how to use it (likely older clinicians), clinicians from orientations that do not emphasize measurement</td>
<td>N/Aa</td>
</tr>
<tr>
<td><strong>Control beliefs</strong></td>
<td></td>
</tr>
<tr>
<td>Please list any factors or circumstances that would make it easy or enable you to use data about clients’ social media use in therapy sessions in the next month.</td>
<td></td>
</tr>
<tr>
<td>Easy-to-use interface that logs all the information on the patient unobtrusively</td>
<td>User-friendly</td>
</tr>
<tr>
<td>Electronic platform, automated data collection and reminders</td>
<td>Convenience</td>
</tr>
<tr>
<td>Easily downloadable app(s), clear directions on how to use it in sessions with clients</td>
<td>Apps</td>
</tr>
<tr>
<td><strong>Discussion</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Principal Findings</strong></td>
<td></td>
</tr>
<tr>
<td>This qualitative study provides new knowledge on patient and therapist perspectives regarding the use of social media in mental health therapy. The interviews captured patients’ and therapists’ current use(s) of social media in mental health therapy and that both patients and therapists initiated its use and discussion in prior sessions. Of note, the individuals interviewed expressed comfort reviewing and discussing social media data. They identified several advantages and disadvantages of sharing social media data in mental health therapy (Table 2), underscore the utility of and the potential concerns associated with integrating elements of our digital lives into mental health therapy in a world that increasingly relies on digital technologies. To our knowledge, this is the first account of patients’ experiences and perspectives on this novel data source. Patients highlighted how social media posts provide objective evidence</td>
<td></td>
</tr>
</tbody>
</table>
of their social lives or networks. They discussed previous experiences of self-curation of posts and questioned if they want an algorithm selecting social media posts to share. Throughout the interviews, a clear understanding of how the data are generated and shared was of great importance. Previous research has captured individuals’ willingness to share digital data for health research [3,22]. However, researchers have yet to devise a way to provide comprehensive feedback or snapshots to better inform patients, let alone their health care provider.

Despite already using social media data in therapy, both patients and therapists questioned the utility of certain social media platforms. They noted that social media posts may not fully reflect their true thoughts or feelings. This speaks to a growing trend among Instagram users with 2 types of accounts: “finsta” accounts, on which users post less polished photos of themselves and “rinsta” accounts, which include more authentic posts [23]. “Finsta” and “rinsta” accounts highlight how social media posts may not always accurately portray individuals’ experiences. However, there might be clinical relevance if a patient has a “rinsta” account.

The desire for objectivity was noted throughout the interviews. Although the interview questions specifically asked about social media data, therapists discussed the incorporation of objective digital data such as the number of steps taken per day or hours slept per night and highlighted them as variables that can contribute to one’s mental health and well-being. Although our interviews did not specifically capture the use of digital data such as steps and screen time, Di Matteo and colleagues [24] reported general acceptance sharing this type of data through their interviews with new patients referred to a tertiary care mood and anxiety disorder clinic. Since Di Matteo et al.’s work [24] and our interviews were both hypothetical scenarios, further research is needed to assess patient and therapist comfort with sharing social media data in the context of therapy sessions.

Lastly, when discussing how to systematically incorporate social media data into sessions, both patients and therapists stressed the importance of person-centered design. As demonstrated by Yoo and colleagues [25], who conducted co-design workshops to build a social media tool for therapists, the data must map onto the end users’ expectations, both in terms of analyses and organization, to avoid being “another layer of noise” [25]. Therapists in this sample questioned how social media data would fit in their workflow (Table 2 and 3). Despite these concerns, overall, our interviews highlight how sharing social media insights with the patient and therapist could also be clinically relevant and informative. Further work is needed to explore and test how to systematically collect social media and present it to patients and their therapists in a user-friendly format for use in mental health therapy.

**Implications and Recommendations**

To our knowledge, this is the first study to capture the patient perspective on their experience sharing and discussing social media data in mental health therapy. The participants interviewed provided critical insights that have yet to be characterized. Our findings uniquely underscore the importance of patient autonomy on what and when to share social media data. As the study sample included mostly female, White, college-educated patients and female, White, and CBT therapists, future research should be conducted with more diverse samples in terms of gender, race/ethnicity, therapeutic orientation, and educational attainment. Additional and special attention is also needed to explore how social media is used and discussed in other cultural contexts.

Our interviews were conducted in 2018 and 2019 prior to the COVID-19 pandemic and may not reflect current norms and beliefs of social media data sharing in the therapeutic encounter. During the COVID-19 pandemic and recovery phases, there has been a dramatic increase in individuals with depressive and anxiety symptoms seeking mental health care [26]. Further research is warranted to capture how social media data are shared in virtual sessions via videoconference or telephone call. We also recommend additional research on the use of digital data such as smartphone metadata as viable data sources in therapy [27]. The combination of social media and digital data could enhance tailored treatment plans and impact therapeutic alliance, a cooperative working relationship between client and therapist, often seen as an essential aspect of successful therapy [28]. Since previous research found that therapeutic alliance is maintained and even enhanced with the introduction of digital mental health interventions [29,30], further research in this area is needed.

Lastly, there are clear educational, practical, and policy implications. As detailed in the American Psychological Association Guidelines for the Optimal Use of Social Media in Professional Psychological Practice (October 2021) [31] and indicated by our findings, mental health therapists should be encouraged to undergo specific educational training on how to safely and ethically use social media data in therapy. The training should encompass core elements of ethics, informed consent, comfort with technology, social media trends per population segment (eg, age, gender, race, ethnicity, sexual identity, language, culture), and how to integrate these new data sources into their workflow. Clinics and practices would need to reserve additional time, personnel, and technology infrastructure to support this training. Noel and colleagues [11] recommend a technology specialist, a new type of health care worker who identifies and reviews electronic resources that may support a client’s specific recovery goals. Data infrastructure and protections are critically important and require special attention. Furthermore, if social media data were incorporated into therapy and demonstrated to improve patient outcomes and reduce costs to the clinic and patient, national/state policies and insurance companies could modify current plans and coverage.

**Limitations**

This study has several limitations. First, the study sample was small and largely homogeneous with respect to sociodemographic data. All participants were recruited from a convenience sample in 1 large metropolitan region in northeastern United States. It is possible that the results from this study do not apply to other population segments or geographic regions. As we advertised for this study online, it is possible that study participants were drawn to the study because of their prior experience with social media in mental health therapy. We were unable to audio record patient...
participant interviews. As such, our transcripts were not as robust for patient participants as they were for therapist participants. Our interviews were conducted with patients and therapists separately; future research could interview patient and therapist dyads for further insights. With respect to analyses, there are limitations to thematic coding as a methodological approach, such as inferences made from a small study sample size and coding at the phrase level, which may not fully capture the participants’ intended meaning. Furthermore, we were unable to use qualitative data analysis software such as NVivo. Interviews were conducted prior to the COVID-19 pandemic. Accordingly, findings may not fully reflect the current state of mental health delivery in the United States, as most mental health therapy is now delivered virtually via videoconference [32]. Despite these limitations, findings from this pilot study can inform social media use practices and norms in mental health therapy.

Conclusions
In this study, patient and therapist interviews provide important insights on the current utilization of social media in mental health therapy, including the advantages and disadvantages of social media use in such contexts. Our findings highlight that social media data used in therapy provide convenient, objective information that is user-friendly and can promote rapport between the patient and the therapist. However, the use of social media data in mental health therapy is also perceived as nonreflective, ethically ambiguous, and potentially nongeneralizable. Future research is needed to explore and test how to systematically collect social media and present it to patients and their therapists in a user-friendly format for use in mental health therapy.

Acknowledgments
We would like to acknowledge and thank Andrea Ngan and Jessica Diaz who assisted with the interviews and interview transcriptions.

Conflicts of Interest
None declared.

References


28. Southwick et al. SMUQ: social media use questionnaire. JMIR Form Res 2022 | vol. 6 | iss. 7 | e32103 | p.85 https://formative.jmir.org/2022/7/e32103

Abbreviations

CBT: cognitive behavioral therapy
SMUQ: social media use questionnaire
A German Smartphone-Based Self-management Tool for Psoriasis: Community-Driven Development and Evaluation of Quality-of-Life Effects

Lea C Brandl1,2, MSc; Claudia Liebram3; Wendelin Schramm4, Prof Dr med; Monika Pobiruchin2,4, Dr sc hum

1Institute of Telematics, University of Lübeck, Lübeck, Germany
2Consumer Health Informatics SIG, German Association for Medical Informatics, Biometry & Epidemiology (GMDS e.V.), Cologne, Germany
3Psoriasis-Netz, Berlin, Germany
4GECKO Institute for Medicine, Informatics & Economics, Heilbronn University, Heilbronn, Germany

Corresponding Author:
Lea C Brandl, MSc
Institute of Telematics
University of Lübeck
Ratzeburger Allee 160
Lübeck, 23562
Germany
Phone: 49 451 3101 6401
Email: brandl@itm.uni-luebeck.de

Abstract

Background: Psoriasis is a chronic disease characterized by inflammation, increased scaling, itching, and other symptoms. Psoriasis is not contagious, but patients have often felt shunned. Therefore, in addition to psoriasis symptoms, stress, anxiety, and depression can also affect quality of life (QoL). Surveys show that only a quarter of patients are satisfied with the success of their therapy. However, in addition to medical therapy, self-management can also make it easier to deal with chronic diseases like psoriasis.

Objective: The aim of this project was to develop a smartphone-based self-management tool (SMT) specifically for patients with psoriasis using a community-driven process. The impact of the SMT on QoL as well as its acceptance and usability were evaluated.

Methods: In collaboration with an internet-based self-help community, 2 user surveys were conducted to determine the requirements for a smartphone-based SMT. The surveys consisted of semistructured questionnaires asking for desired features in an SMT for psoriasis. A pilot study was conducted to evaluate QoL, acceptance, and usability. Community users were recruited to use the app for 21 days and complete the Dermatology Life Quality Index (DLQI) questionnaire at the beginning (T0) and end (T1). Afterward, participants were asked to complete another questionnaire on usability and ease of use.

Results: SMT requirements were collected from 97 members of an internet-based community. The SMT was built as a progressive web app that communicates with a server back end and an Angular web app for content management. The app was used by 15 participants who also provided qualitative feedback, and 10 participants answered all questionnaires. The average DLQI score was 7.1 (SD 6.2) at T0 and 6.9 (SD 6.6) at T1. The minimal required sample size of 27 was not reached.

Conclusions: The high degree of community participation in the development process and the responses during the requirement engineering process indicated that there is a general need for an independently developed SMT for patients with psoriasis. However, the feedback received after app use shows that the SMT does not meet the needs of the community. It can be concluded that a more customizable app is needed. The focus and needs of the users were very heterogeneous. Similar developments and research could benefit from the findings of this project.

(JMIR Form Res 2022;6(7):e32593) doi:10.2196/32593

KEYWORDS
psoriasis; self-management; mobile apps; quality of life; mobile phones; smartphones
Introduction

Overview
Psoriasis is a chronic skin disease that manifests itself with inflammation, increased scaling, itching, and other symptoms [1]. Flares can be aggravated by stress, medical drugs, or infectious diseases [2-4]. Psoriasis affects approximately 2% to 3% of the German population [1,5]. In addition to psoriatic symptoms, stress, anxiety, and depression can affect quality of life (QoL) [6-8]. Compared to healthy individuals, anxiety disorders were identified more frequently in patients with psoriasis [4,9]. It is not possible to cure psoriasis, but symptoms could be mitigated with medical drugs, special therapies, or complementary methods [10,11]. Nevertheless, approximately 30% of patients do not consult a physician [12,13]. Interviews reveal that a quarter of patients are dissatisfied with the success of their treatment [14,15]. Problems regarding drug therapies are present in up to 40% of patients [15,16]. To increase patient satisfaction and adherence, evidence-based decision guidance for psoriasis therapy is available as an S3 guideline [17]. However, in addition to medical therapy, self-management can facilitate the management of chronic diseases such as psoriasis [18-20]. In comparison to other chronic diseases, there are less evidence-based self-management tools (SMTs) for patients with psoriasis [20].

Related Work
Self-help and self-management can be effective tools for patients to better cope with their (chronic) disease [18-20]. Traditionally, self-help involves in-person group meetings or counseling. In recent years, a shift toward mobile apps, information websites, and web-based communities can be observed. Self-help can also include aspects of citizen science approaches. For example, self-help groups and other organizations collect and organize information on people with a disease [21]. Internet-based communities are often used by patient experts who share information and generate hypotheses. This is done based on shared experiences. Modern technologies can facilitate new approaches to citizen science projects [22].

Citizen science also includes the involvement of patients in the development process of SMTs. Some examples are given in the following paragraphs. Safdari et al [23] conducted a study on the requirements for a self-management app for patients with psoriasis. In a requirements analysis, 100 patients provided information about their requirements for educational information and lifestyle management, among other features. This included information on the disease on the one hand and factors such as physical activity, nutrition, or stress management on the other hand.

Trettin et al [24,25] have developed an app for Danish patients with psoriasis treated with biologics. In the development process, patients were interviewed, and various workshops and a prototype test were carried out. In the app, vital signs, and the Dermatology Life Quality Index (DLQI) [26] can be recorded in preparation for video or telephone consultations. Patients, doctors, and nurses have reported that consultations are more structured, and patients feel safe [24,25].

Aims of the Study
Currently, there are few SMTs in German app stores and no German-speaking apps (co-) developed by self-help organizations (ie, the patients themselves). Therefore, the 2 major objectives of this study are to (1) present an SMT especially developed by and for patients with psoriasis and (2) evaluate the app’s impact on QoL as well as its acceptance and usability.

Methods

Recruitment and Requirement Engineering Phase
In Germany, there are 2 main organizations for psoriasis self-help. The larger one is the Deutscher Psoriasis Bund e.V. (German Psoriasis Association) and the smaller one is the Psoriasis Selbsthilfe Arbeitsgemeinschaft e.V. (Psoriasis Self-Help Association, PSOAG) [27,28]. The latter offers self-help on its website. The internet-based community consists of approximately 28,000 users (as of November 2019) and concerns itself with topics such as therapies or nutrition. There are various expert forums and groups that can be joined. Knowledge articles are also published [29]. The project was initiated with PSOAG members. The community collected requirements of a (potential) smartphone-based SMT during 2 online surveys (published via the web-based forum).

The first survey started on February 2, 2019, and was closed on March 15, 2019. It consisted of a semistructured questionnaire that asked about the desired features of an SMT for psoriasis. A translated version of this questionnaire can be found in Multimedia Appendix 1. After completion of the first survey, the responses were read and clustered to capture different categories of functionalities. In the next step, categories were checked for technical feasibility and focus on SMTs.

The second survey that incorporated the community feedback from the first round was open from June 25, 2019, to August 31, 2019. During this phase, 11 mock-ups of the SMT were presented and discussed with the community members. The mock-ups were built with the Balsamiq Mockup software (version 3.5.17; Balsamiq Studios, LLC) [30]. Changes needed in the design of the app and the expansion of certain functionalities were again checked for technical feasibility and incorporated into the final app requirements.

After the requirements were determined, the SMT development started. The system was implemented by one of the authors using software components previously developed [31] as blueprints. The app development and the pilot study were carried out as part of a master’s thesis at Heilbronn University.

SMT Evaluation

QoL Assessment
Numerous instruments of QoL assessment exist [32,33]. Fitzpatrick et al [34] developed a list of criteria to select the most appropriate measurement instrument. According to these criteria, we decided to use the DLQI by Finlay et al [26] for this study. The DLQI consists of 10 questions that cover the dimensions of symptoms and daily activity, leisure, work or...
school, personal relationships, and therapy [26]. The average response time is 2 minutes [35]. The score of the DLQI ranges from 0 to 30. The larger the value, the worse the QoL [26]. The minimal clinically important difference is 4 score points [36]. A license for the use of the German Translation DLQI was applied for and was provided by Cardiff University [35].

Use and Usability
Considering usability and further use, an additional questionnaire was introduced. It included questions about handling, use, further use, and the severity of psoriasis, and were self-reported and/or measured by the Psoriasis Area and Severity Index (PASI) [37]. This additional questionnaire can be found in Multimedia Appendix 2.

Recruitment for QoL Assessment
Participants for the QoL assessment were recruited through various channels:

1. A call for participation was given via the PSOAG website and forum. The call provided the opportunity to contact the study team directly or to access a study information website.
2. Information flyers were sent to dermatology clinics (13 rehabilitation clinics and 11 university clinics), the Professional Association of German Dermatologists, and the Psoriasis Association.
3. Social media channels were used to actively promote the study (eg, regular tweets, posts on Instagram, and the creation of a Facebook page about the study).

For inclusion in the study, the participants had to have psoriasis and be at least 18 years old.

Study Design
As the study measured the change in QoL, it was necessary to record it at a minimum of 2 time points ($T_0$ and $T_1$). The DLQI measures QoL in relation to the last 7 days. A break of at least 7 days is recommended, and frequent questioning is discouraged; otherwise, participants may remember their previous response. Based on these requirements, the intervention duration was set at 21 days.

After 21 days had elapsed, the second DLQI and the extra questionnaire on usability and further use of the SMT were completed. Then, the participants dropped out of the study.

Statistical Analysis
Based on the German-specific data provided by Lesner et al [7] and the minimal clinical effect of 4 points, a sample size of 27 participants is calculated to reach $\alpha = .05$ and $\beta = .2$ [38]. Data were analyzed with the statistical software MATLAB (version 2019b) [39].

Ethical Approval
No ethical approval was obtained for the conducted study. The local ethics committee of Heilbronn University is only an advisory board (see §1(1) [40]), and it does not provide formal ethical approval. This was a community-led project involving interested persons with psoriasis who enrolled voluntarily and by general invitation in the development and evaluation of an app. These were not patients treated by the authors. Treatment changes or any interaction with the medical care delivery team were not included in the objectives of this study.

As the data collected were worthy of protection, a high scientific standard was applied here. All participants were informed about the study details and received written information in accordance with the Declaration of Helsinki [41]. This information was available on the study information website and could be downloaded for offline reading (see Multimedia Appendix 3).

All participants signed an informed consent form (see Multimedia Appendix 4), which could be revoked at any time. In this case, all documents that could still be assigned were destroyed. To ensure a high level of data security, the questionnaires were recorded under a pseudonym generated by the participants themselves.

Results

Requirement Engineering and SMT
More than 90 community members participated in the survey (N=97). They determined the scope of the SMT’s features, created the medical content, and curated it.

During the first user survey, the requirements for the SMT were collected. Nearly 90% (87/97) of the participants indicated willingness to test the SMT. The suggestions of the community about the desired features were evaluated for their feasibility and purpose of use. Four categories emerged during the analysis of the community responses: (1) communication, (2) drug management, (3) tracking of complementary methods, and (4) rate/score doctors. Selected examples of the feedback from the SMT users are presented in Table 1. Furthermore, mock-ups (ie, drafts of the user interface) were created. These were presented to the community and the community members could provide their comments.

The final SMT framework consists of three components, as shown in Figure 1: (1) an Angular web application for the management of the SMT content used by the editorial team, (2) a server back end built with Java Spring Boot and Spring Security, and (3) the SMT for the study participants as an Angular application. The SMT app was implemented as a progressive web app (PWA) to be available for all types of smartphones.

Both web components were built with the Angular Framework (Google; version 7.2.15) [42]. The server back end was built with Java Spring Boot (The Spring Team; version 2.1.5) [43] and Java (Oracle Corporation; version 11) [44].

One feature focus of the SMT is to suggest interaction-free complementary measures to the user, with which physical complaints, such as itching, skin blisters, and dry skin, can be alleviated. Patients can document their psoriasis type, medical drugs, and complaints each day. Based on these data, the SMT suggests complementary methods that are free of interactions. The complementary measures were suggested by the community itself and were reviewed by the forum’s editorial team. In total,
55 complementary measures were found, described, and incorporated into the knowledge base of the SMT (see component 1 in Figure 1). The complete list is available in Multimedia Appendix 5. The knowledge was embedded in a PostgreSQL database (PostgreSQL Global Development Group; version 10.9) [45], as shown in Figure 1.

Another aspect of the app was the documentation of psoriasis and the possibility to view the severity of symptoms as they progressed and in relation to the complementary methods that were considered. Moreover, individual body parts could be documented in more detail with the help of photos and text. Figure 2 shows representative screenshots.

A comprehensive data protection concept was developed. All the app data were stored exclusively on the user’s own smartphone.

Table 1. Results of the first requirement analysis involving the web-based communitya.

<table>
<thead>
<tr>
<th>Feedback from the community</th>
<th>Category</th>
<th>Realization in the SMTb</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Input of laboratory values and their graphical representation in order to be able to observe their development (eg, CRP, leukocytes, lymphocytes)”</td>
<td>2</td>
<td>_d</td>
</tr>
<tr>
<td>“Alarm clock to remind of injection days, medication intake, doctor’s appointments”</td>
<td>2</td>
<td>—</td>
</tr>
<tr>
<td>“One should be able to print out what has been written”</td>
<td>3</td>
<td>✓</td>
</tr>
<tr>
<td>“Being up to date on care for scalp psoriasis - nutrition tips, new findings”</td>
<td>3</td>
<td>+f</td>
</tr>
<tr>
<td>“Diet plan, natural remedies”</td>
<td>3</td>
<td>+</td>
</tr>
<tr>
<td>“Exchange with other patients (ie, link to the forum)”</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>“Notifications about new findings regarding the therapy used (side effects, new variants)”</td>
<td>2</td>
<td>—</td>
</tr>
<tr>
<td>“All [measures] that improve my quality of life, take away pain, and build me up”</td>
<td>3</td>
<td>+</td>
</tr>
<tr>
<td>“It could remind me to take medication”</td>
<td>2</td>
<td>—</td>
</tr>
<tr>
<td>“Study situation and results on complementary therapy methods would be good”</td>
<td>3</td>
<td>✓</td>
</tr>
<tr>
<td>“Some kind of daily checklist to check off to-dos related to (psoriasis) would be great. That way you can see correlations, if necessary (eg, if you forgot a supplement or only applied cream once a day instead of twice).”</td>
<td>3</td>
<td>+</td>
</tr>
<tr>
<td>“Could you name and comment on current dermatologist (recommend, yes/no and why)”</td>
<td>4</td>
<td>—</td>
</tr>
<tr>
<td>“That you can print out the diary would be good to present to the doctor who treated you!”</td>
<td>3</td>
<td>✓</td>
</tr>
<tr>
<td>“A list of my current medications as well as a list of medications I have tried but no longer use (and why!)”</td>
<td>2, 3</td>
<td>+</td>
</tr>
<tr>
<td>“A chat or a link to a forum where users can exchange information directly. Because I see the main problem with the app as being that something different helps everyone. It’s so hard to generalize.”</td>
<td>1</td>
<td>+</td>
</tr>
<tr>
<td>“Calendar in which one can enter appointments for the doctor, taking one’s medication or other things.”</td>
<td>1, 2</td>
<td>—</td>
</tr>
<tr>
<td>“Logging what you have eaten and then automated evaluation of whether patterns are recognizable, how nutrition affects you. The same with stress/well-being.”</td>
<td>3</td>
<td>✓</td>
</tr>
<tr>
<td>“Maybe refer to competent doctors, dermatologists and rheumatologists in the respective federal states of the people concerned.”</td>
<td>4</td>
<td>—</td>
</tr>
<tr>
<td>“To find other patients with psoriasis nearby and find experts or good doctors from the patient’s point of view.”</td>
<td>1, 4</td>
<td>—</td>
</tr>
<tr>
<td>“A digital (medication plan) for all medications, not only for psoriasis treatment; in my case, for example, this was accompanied by a CHD disease.”</td>
<td>2</td>
<td>—</td>
</tr>
</tbody>
</table>

aAll statements were translated from German. Additions or replacements for better understanding are denoted in square brackets.

bSMT: self-management tool.

cCRP: c-reactive protein.

dThe function is not implemented.

eThe function is fully implemented.

fThe function is partially implemented.

gCHD: congenital heart defects.
Figure 1. System architecture. The 2 Angular components are shown on the left. These are provided on a web server that communicates with the back-end server. A Spring Security module protects the back end and the connected database from security attacks. The back-end component is shown on the right. DB: database; SMT: self-management tool; SQL: structured query language.

Figure 2. Screenshots of the smartphone app. Left to right: (A) Rating and sorting physical complaints. Complaints can be added via the “plus” sign in the footer. (B) Choosing complementary methods, including reducing alcohol, taking a bath with bath salts, and relaxing. (C) The course of complaints. By clicking on one of the dots in the chart, the complementary methods used during this day are shown. (D) Documentation of the front of the body. A click on the body adds a region of interest. Photos and text could be documented for each region of interest.

SMT Evaluation

QoL Analysis
During the call for participation, 29 persons showed interest in participation, with 21 of them providing consent to the processing of their data. Of these, 18 submitted the initial questionnaires. All persons who returned the first questionnaire are counted as participants. Participants who revoked consent (3 participants) or did not return the second questionnaire (5 participants) are counted as dropouts. Only 10 participants returned both questionnaires that were subsequently analyzed.

In total, 15 valid preintervention and 10 postintervention QoL questionnaires, and questionnaires about acceptance and usability were collected and transcribed into a CSV format. Additionally, 6 participants did not return any questionnaires but gave consent.

The mean DLQI for the preintervention questionnaire was 7.1 (SD 6.2) score points and 6.9 (SD 6.6) score points for the postintervention questionnaire. A lower score after the intervention corresponds to an increased QoL. However, the reduction of the DLQI by 0.2 score points does not correspond to a clinically relevant effect.

Use and Usability
The survey on difficulties in using the app shows that users did face challenges (see Table 2).

The results showed that 4 participants had no difficulties in using the app, and 3 participants found it quite or very difficult to use. Moreover, 3 other participants had slight difficulty in using the app. This was also reflected in the feedback from the participants. When returning the second questionnaire or withdrawing from the study, there was additional qualitative feedback on the SMT given by various participants, which is documented inTextbox 1.

The participants were also asked about their continued use of the app. None of the participants would continue to use the app daily. Only 2 participants continued to use the app regularly. 1 participant used it as needed, and 6 participants did not continue to use the app at all. In 1 questionnaire, this question remained unanswered.
### Table 2. Study results.

<table>
<thead>
<tr>
<th>Participant</th>
<th>DLQI&lt;sub&gt;T0&lt;/sub&gt;</th>
<th>DLQI&lt;sub&gt;T1&lt;/sub&gt;</th>
<th>Usage period (days)&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Difficulty&lt;sup&gt;e&lt;/sup&gt;</th>
<th>Further use&lt;sup&gt;f&lt;/sup&gt;</th>
<th>Severity level&lt;sup&gt;g&lt;/sup&gt;</th>
<th>PASI&lt;sup&gt;h&lt;/sup&gt; score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>A little</td>
<td>Not at all</td>
<td>Mild</td>
<td>Not replied</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>3</td>
<td>12</td>
<td>Fairly</td>
<td>Not replied</td>
<td>Heavy</td>
<td>Not replied</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>A lot</td>
<td>Not at all</td>
<td>Heavy</td>
<td>Not replied</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>19</td>
<td>6</td>
<td>A little</td>
<td>Not at all</td>
<td>Mild</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>2</td>
<td>21</td>
<td>Not at all</td>
<td>Fairly</td>
<td>Not replied</td>
<td>Not replied</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>Not at all</td>
<td>Not at all</td>
<td>Moderate</td>
<td>Not replied</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>4</td>
<td>8</td>
<td>Not at all</td>
<td>Not at all</td>
<td>Moderate</td>
<td>&lt;3</td>
</tr>
<tr>
<td>8</td>
<td>20</td>
<td>17</td>
<td>5</td>
<td>A little</td>
<td>As needed</td>
<td>Moderate</td>
<td>Not replied</td>
</tr>
<tr>
<td>9</td>
<td>15</td>
<td>9</td>
<td>5</td>
<td>Not at all</td>
<td>Regularly</td>
<td>Heavy</td>
<td>Not replied</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>12</td>
<td>3</td>
<td>Not at all</td>
<td>Not at all</td>
<td>Moderate</td>
<td>Not replied</td>
</tr>
<tr>
<td>11</td>
<td>15</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>12</td>
<td>15</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>14</td>
<td>10</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

<sup>a</sup>Average values for participants 1 to 10: DLQI<sub>T0</sub>, 7.1, DLQI<sub>T1</sub>, 6.9, and usage period (days), 7.1. Average DLQI<sub>T0</sub> value for participants 1 to 15 was 8.3.

<sup>b</sup>DLQI<sub>T0</sub>: Dermatology Life Quality Index before using the app.

<sup>c</sup>DLQI<sub>T1</sub>: Dermatology Life Quality Index after using the app.

<sup>d</sup>Indicates the number of days the app was used.

<sup>e</sup>Represents the difficulties experienced when using the app.

<sup>f</sup>Represents further use of the app.

<sup>g</sup>Shows the subjective perception of the severity of psoriasis.

<sup>h</sup>PASI: Psoriasis Area and Severity Index.

<sup>i</sup>Indicates participants who only completed the first questionnaire but did not withdraw their agreement.
Textbox 1. Qualitative feedback of participants. All statements were translated from German. Additions for better understanding are shown in square brackets. Detailed information on date, time, and medication are omitted because of privacy reasons.

**Person A**
- “I have psoriatic arthritis and unfortunately the app does not relate to that.”
- “And unfortunately, I also have to say that I find the handling very complicated.”
- “Basically, I find the idea good, but so [the app] is unfortunately not further usable for me.”

**Person B**
- “It worked great.”
- “The “problem” is the photos. I live alone and find it difficult to photograph the back. But that is my only issue at the moment.”

**Person C**
- “I have tried a few apps, but they were all not applicable. Maybe theirs is a little better and more supportive.”

**Person D**
- “Also, I found it inconvenient and difficult to use at times because there was no explanation whatsoever.”

**Person E**
- “The “app” is not a native iOS or Android app, but a web application. I prefer here, especially for my sensitive health data, an app running locally on the iPhone/iPad, where the data is local on the device or encrypted in my own or the iCloud (optional).”
- “Using the app in parallel from multiple devices (smartphone, tablet, smartwatch) would be of great benefit.”
- “Such an app would also have to be individually “customizable” for me: Creation of own “therapies”. For the individual therapies, storage of more details (eg, for light therapy, the duration of the respective irradiation or the set Joule dose), for ointments, the name, PZN, etc.”
- “Usability would also have to be significantly optimized. For regular documentation (skin condition, condition, medication), the [documentation] must happen as quickly and easily as possible.”

**Person F**
- “The app is very interesting and helpful for people who don’t have this background knowledge.”
- “The documentation via photos, I think is very good, I could have used that a lot from ***.”

**Person G**
- “I only used the app for a very short time, as I perceived filling it out as annoying.”
- “However, I think small changes to the app could fix this for the most part.”

**Person H**
- “Since I don’t do therapy other than *** and *** and don’t yet know what things help me, it would have also become difficult for me to use it meaningfully.”
- “I’m just looking more the other way around for a template where I can document what I’ve done and eaten (how much sleep, how much sun, etc) to figure out what factors are negatively impacting me.”

**Statistical Analysis**
The targeted sample size (N=27) was not achieved. Therefore, the testing of the hypotheses was disregarded.

**Discussion**

**Requirements Engineering and SMT Development**
A special feature of the development process was the high level of participation. Community members were asked about additional features and could comment on the designs and mock-ups. Unfortunately, it was not possible to implement all the desired functions (see Table 1). Reminders could not be implemented due to the technical restriction of a PWA. Notifications are possible with this design but cannot be individually configured. Ratings of doctors and exchanges with other patients were not implemented, as these functions would overlap with the internet-based forum of the community. Some suggestions of the stakeholders have been partially implemented because the documentation should be in a structured form instead of a free form to avoid incorrect entries and simplify usage.

When examining for possible confounders, the usability aspects stood out. This was reinforced by the feedback from the individual participants. There were difficulties in dealing with
the app. This was partly due to the handling and partly because patients have a greater need to document their illness than that assumed. In particular, the behavior for which therapy is currently being provided and secondary diseases should be documented more precisely and as quickly as possible. The heterogeneity of the users’ feedback shows that the app should be highly customizable. The suggestion to implement the app as a standalone one rather than as a PWA certainly warrants further IT security requirements and offers possibilities for including more features.

The feedback from the community shows a certain ambivalence regarding the privacy aspect. On the one hand, previous experience and feedback on the app show that a high level of data protection is desired. Therefore, all data remain locally stored on the user’s device. On the other hand, the fact that the data cannot be accessed on several devices was perceived as a negative feature. Apparently, there are services that enjoy a higher level of trust regarding data protection than others. Further developments to the app must ensure that users can decide whether they want to use the SMT data on 1 or more devices.

Despite the low number of study participants, the project shows the possibility of using citizen science in biomedical projects and research. The community-supported software development process successfully led to a functional SMT. In particular, the internet-based community provided the list of complementary measures, which were used as textual content in the SMT. Therefore, the SMT users could benefit from the collected patient knowledge about complementary measures without the effort of searching information on the internet.

Limitations

There are 2 major limitations of this project. The first is the low overall sample size and the second is a small observed effect.

For showing a statistically significant improvement of a minimum of 4 score points, 27 participants would have been necessary for the study. This number was not reached despite previous experience about the active internet-based community and high involvement in the requirement engineering process. Only 21 participants could be recruited. Of these, only 10 participants completed all questionnaires. We identified potential reasons for this: (1) failure to reach the target group via social media channels, (2) an admission process that was perceived as too complicated, and (3) the Christmas holidays. One indication of this is an increase in the dropout rate in the month of December. The additional leaflets sent in December could not mitigate the low involvement.

The observed effect of 0.2 score points reported in the results is negligible. The fact that the desired effect with a difference of 4 DLQI points could not be measured can be attributed to different causes. Besides the low effectiveness of the intervention, the intervention period (21 days) may play a role. The reason is that regular use over a long period of time may be necessary to effect changes. Therefore, a measured effect could be considered a placebo effect. In addition, the DLQI was recorded without considering the occurrence of relapses, possible rehabilitation stays, or an existing concomitant disease such as psoriatic arthritis. Comparing the average DLQI score determined in this study with the results of the study by Augustin et al [46] (mean 7.5, SD 6.4 points) reveals that the values do not differ much. Moreover, the variances in the respective samples are similar. This difference of less than 1 score point with the score in this study is noteworthy because Augustin et al included not only patients of dermatology outpatient hospital clinics but also patients of dermatologists in ambulant care. In this study, the participants were recruited via the community, which is why it is unlikely that the sample consists exclusively of dermatology outpatient hospital clinic patients.

Future Directions

Based on the feedback provided by the community and app users, the existing smartphone app should be completely revised. The authors strongly recommend developing such an app using a community-driven process again.

The improvements to the subsequent versions of the SMT can be roughly divided into four areas: (1) therapies including conventional medical therapies, hospital and rehabilitation stays, and complementary measures; (2) psoriasis relapses and triggers that can be analyzed; (3) behavioral patterns or environmental data in everyday life; and (4) QoL, as well as psychological and physical factors measured textually or visually using questionnaires.

To implement SMTs sustainably in care, it should be better integrated into the existing structures of care processes and medical practices. Direct data transmission to practitioners and further development of SMTs as approved medical products are conceivable.

Conclusions

The high participation of the internet-based community in the development process and the response to the first survey (N=97) shows a general need for independently developed SMTs for people with psoriasis. However, the collected feedback shows that the solution presented in the paper does not meet the needs of the community. The authors conclude that a more customizable app is needed.

Acknowledgments

The authors would like to thank all the members of the psoriasis-netz.de community who provided feedback; all the participants of the study for their time, effort, and feedback on the app; the team from Cardiff University for their support in granting the DLQI license; and the staff in Heilbronn University (especially the groups in E112 and E117) for their continuous support throughout the SMT development.
Conflicts of Interest
None declared.

Multimedia Appendix 1
Semistructured German questionnaire and English translation used in the requirement engineering process.
[DOCX File, 16 KB - formative_v6i7e32593_app1.docx]

Multimedia Appendix 2
Original German questionnaire and English translation for usability and acceptance of the self-management tool.
[DOCX File, 26 KB - formative_v6i7e32593_app2.docx]

Multimedia Appendix 3
Study information website with information on the project that can be downloaded as a PDF file.
[PDF File (Adobe PDF File), 1489 KB - formative_v6i7e32593_app3.pdf]

Multimedia Appendix 4
Informed consent of participants.
[PDF File (Adobe PDF File), 1208 KB - formative_v6i7e32593_app4.pdf]

Multimedia Appendix 5
List of complementary measures suggested in the app, considering interactions.
[DOCX File, 17 KB - formative_v6i7e32593_app5.docx]

References


27. Selbsthilfe vor Ort. Deutscher Psoriasis Bund eV. URL: https://www.psoriasis-bund.de/selbsthilfe-vor-ort/ [accessed 2021-07-08]


30. Mockups 3 for Desktop | Balsamiq. Balsamiq Studios, LLC. URL: https://balsamiq.com/wireframes/mockups3fordesktop/ [accessed 2021-07-08]


42. The modern web developer's platform. Angular. URL: https://angular.io/ [accessed 2021-07-08]
43. Webb P. Spring Boot 2.1.5 released. Spring. 2019 May 15. URL: https://spring.io/blog/2019/05/15/spring-boot-2-1-5-released [accessed 2021-07-08]

Abbreviations

- **DLQI**: Dermatology Life Quality Index
- **PASI**: Psoriasis Area and Severity Index
- **PSOAG**: Psoriasis Selbsthilfe Arbeitsgemeinschaft eV (Psoriasis Self-Help Association)
- **PWA**: progressive web app
- **QoL**: quality of Life
- **SMT**: self-management tool

Edited by T Leung; submitted 03.08.21; peer-reviewed by S Gulliver, T van Hal; comments to author 11.10.21; revised version received 08.11.21; accepted 01.06.22; published 07.07.22.

Please cite as:
Brandl LC, Liebram C, Schramm W, Pobiruchin M
A German Smartphone-Based Self-management Tool for Psoriasis: Community-Driven Development and Evaluation of Quality-of-Life Effects
JMIR Form Res 2022;6(7):e32593
URL: https://formative.jmir.org/2022/7/e32593
doi:10.2196/32593
PMID:35797109

©Lea C Brandl, Claudia Liebram, Wendelin Schramm, Monika Pobiruchin. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.07.2022. 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 Users' Experiences With a Quick-Response Chatbot Within a Popular Smoking Cessation Smartphone App: Semistructured Interview Study

Alice Alphonse¹, MSc; Kezia Stewart¹, MSc; Jamie Brown¹, PhD; Olga Perski¹, PhD
Department of Behavioral Science and Health, University College London, London, United Kingdom

Corresponding Author:
Olga Perski, PhD
Department of Behavioral Science and Health
University College London
1-19 Torrington Place
London, WC1E 7HB
United Kingdom
Phone: 44 0207679 ext 1258
Email: olga.perski@ucl.ac.uk

Abstract

Background: Engagement with smartphone apps for smoking cessation tends to be low. Chatbots (ie, software that enables conversations with users) offer a promising means of increasing engagement.

Objective: We aimed to explore smokers’ experiences with a quick-response chatbot (Quit Coach) implemented within a popular smoking cessation app and identify factors that influence users’ engagement with Quit Coach.

Methods: In-depth, one-to-one, semistructured qualitative interviews were conducted with adult, past-year smokers who had voluntarily used Quit Coach in a recent smoking cessation attempt (5/14, 36%) and current smokers who agreed to download and use Quit Coach for a minimum of 2 weeks to support a new cessation attempt (9/14, 64%). Verbal reports were audio recorded, transcribed verbatim, and analyzed within a constructivist theoretical framework using inductive thematic analysis.

Results: A total of 3 high-order themes were generated to capture users’ experiences and engagement with Quit Coach: anthropomorphism of and accountability to Quit Coach (ie, users ascribing human-like characteristics and thoughts to the chatbot, which helped foster a sense of accountability to it), Quit Coach’s interaction style and format (eg, positive and motivational tone of voice and quick and easy-to-complete check-ins), and users’ perceived need for support (ie, chatbot engagement was motivated by seeking distraction from cravings or support to maintain motivation to stay quit).

Conclusions: Anthropomorphism of a quick-response chatbot implemented within a popular smoking cessation app appeared to be enabled by its interaction style and format and users’ perceived need for support, which may have given rise to feelings of accountability and increased engagement.

(JMIR Form Res 2022;6(7):e36869) doi:10.2196/36869

KEYWORDS

chatbot; conversational agent; engagement; smartphone app; smoking cessation; accountability; mobile phone

Introduction

Diseases caused by cigarette smoking are a leading cause of preventable death, killing approximately 8 million people each year globally [1]. Smoking-attributable diseases place significant financial burden on health care systems, with costs estimated to be approximately 5.7% of the total annual global health care expenditure [2]. Therefore, improved behavioral or pharmacological smoking cessation support is a priority for individuals, public health bodies, and governments. However, in-person smoking cessation services are challenged by scalability and face substantial funding cuts across many countries [3], especially after many had to offer only remote services owing to the COVID-19 pandemic [4]. With growing internet access and smartphone ownership, digital interventions (including smartphone apps) provide a low-cost means of scaling up the delivery and optimizing the reach of evidence-based smoking cessation support [5]. However, the available smoking cessation apps tend to generate low average levels of user engagement [6,7]—although estimates vary between apps
Chatbots are a relatively new addition in the health care domain, with recent systematic reviews identifying only a handful of studies of chatbots for improving mental health [23] and increasing physical activity and healthy diets [24]. Within the substance use and smoking cessation domains, a few early single-arm and 2-arm randomized studies have yielded promising results [25-28]. For example, a chatbot incorporating the principles of motivational interviewing—designed specifically to support smokers who are unmotivated to stop—tested positively in an early user-testing study [25]. An adapted version of the cognitive behavioral therapy–informed chatbot, Woebot, for people who use addictive substances was found to be acceptable to deliver, engaging, and associated with improvements in mental health and substance use outcomes in a single-arm study [28]. We found that the addition of a supportive, quick-response chatbot to a popular smoking cessation app more than doubled the user engagement and improved short-term quit success in a large, 2-arm, experimental study [27]. However, the available single-arm and 2-arm quantitative studies have not focused on the potential mechanisms underpinning this increased engagement (eg, owing to limited data collection). Qualitative studies of users’ experiences with the relational agent, Replika, have found that such companion chatbots can mimic human interaction, with users perceiving their relationships with the designated bot as rewarding [20,29]. However, qualitative investigations of users’ experiences with chatbots designed specifically to support smoking cessation are lacking. Therefore, this qualitative study aimed to address the following research questions:

1. What are smokers’ experiences with a quick-response chatbot (Quit Coach) implemented within a popular smartphone app?
2. What are the factors that influence users’ engagement with Quit Coach?

Methods

Study Design

The Consolidated Criteria for Reporting Qualitative Research checklist was used in the design and reporting of this study [30]. Semistructured, one-to-one interviews were conducted.

Theoretical Framework

A constructivist theoretical framework was used to inform data collection and analysis [31]. This theoretical approach was selected because constructivism recognizes the active role of the researcher in the generation and interpretation of qualitative data.

Participants

For pragmatic purposes, participants were recruited across 2 periods: June 2020 to August 2020 (led by KS and OP) and April 2021 to August 2021 (led by AA and OP). Owing to the COVID-19 pandemic, it was challenging to recruit as planned during the summer of 2020. Therefore, we continued the recruitment in 2021. The project team decided that it would be useful to recruit participants from 2 different subgroups (ie, past-year smokers who had voluntarily used Quit Coach in a recent smoking cessation attempt and current smokers who...
agreed to download and use Quit Coach for a minimum of 2 weeks to support a new cessation attempt) as a form of triangulation [32]. We reasoned that such triangulation of results when varying the eligibility criteria (rather than the methods) would either help to validate the results (eg, if smokers who did not self-select to download and use the Smoke Free app had similar experiences with Quit Coach as those who had voluntarily used the app) or highlight different experiences owing to smoking status or treatment-seeking behavior.

Participants recruited in 2020 were eligible to participate if they (1) were aged ≥18 years, (2) were fluent English speakers based in the United Kingdom, (3) were past-year smokers and had used the pro (ie, paid) version of the Smoke Free app (ie, the app version that included Quit Coach) for at least two weeks, and (4) had interacted with Quit Coach at least once during the 2-week period.

Participants recruited in 2021 were eligible to participate if they (1) were aged ≥18 years; (2) were fluent or highly competent English speakers, with no restrictions on geography; (3) were current cigarette smokers; (4) were willing to make a quit attempt within 1 week from initial contact with the researchers and use Quit Coach for at least two weeks; and (5) owned a smartphone.

All the participants used the pro version of the Smoke Free app for a period of at least 2 weeks before participating in the semistructured interviews. We expected this time window to be sufficient for enabling detailed conversation about participants’ chatbot experiences.

**Sampling**

Participants recruited in 2020 were approached through advertisements (unpaid) shared on social media platforms (ie, Facebook and Twitter) and through a mailing list of Smoke Free app users. The recruitment materials stated that Smoke Free users were invited to participate in a web-based interview about their experiences with the app (good or bad), with particular focus on the Quit Coach feature. Participants were incentivized to win 1 of 5 gift vouchers worth £20 (approximately US $24).

Participants recruited in 2021 were approached through advertisements (unpaid) shared on social media platforms (ie, LinkedIn, Facebook, and Instagram), directly through the researchers’ networks (ie, WhatsApp, email, SMS text messages, and flyers), and through professional web-based recruitment platforms (ie, Prolific and Call for Participants). The recruitment materials stated that smokers interested in making a quit attempt with the use of a smartphone app were invited to participate in a web-based interview about their experiences with the app (good or bad), with particular focus on its chatbot feature. Participants received a gift voucher worth £10 (approximately US $12) after completing the interview.

Participants were recruited in batches of 4 to 5 participants each until theoretical saturation was judged to have occurred (ie, a point in the data collection process when no new information alters the identified themes) [33]. Preliminary data analysis was conducted by KS and subsequently by AA after each batch of 4 to 5 participants, to determine whether additional participants were needed.

**Measures**

**Eligibility and Sample Characteristics**

Data were collected to determine eligibility and characterize the sample based on (1) age; (2) gender (female, male, or in another way); (3) country of residence; (4) whether they were fluent or highly competent English speakers (yes or no); (5) time to first cigarette (<5, 6-30, 31-60, or >60 minutes or not applicable); (6) cigarettes smoked per day (<10, 11-20, 21-30, ≥31, or not applicable); and (7) motivation to stop, measured with the validated Motivation to Stop Scale [34].

Participants recruited in 2020 were asked the questions mentioned previously and to provide additional information on the following: (1) job type (manual or nonmanual); (2) smoking status (“I smoke cigarettes [including hand-rolled] every day”; “I smoke cigarettes [including hand-rolled], but not every day”; “I don’t smoke cigarettes at all, but I do smoke tobacco of some kind [eg, pipe, cigar or shisha]”; “I have stopped smoking completely in the last year”; “I stopped smoking completely more than a year ago”; or “I have never been a smoker [i.e. smoked for a year or more]”); and (3) self-reported use of Quit Coach (none at all, a little, moderately, a lot, or extremely).

Participants recruited in 2021 were asked the questions mentioned previously and to provide additional information on the following: (1) the number of past-year quit attempts and (2) whether they had ever used any app-based support to help stop smoking (and if so, the name of the app).

**Interview Topic Guide**

The topic guide was informed by the Model of Supportive Accountability [22] to address specific theoretical concepts (eg, information quality, reliability, and accountability) and split into 3 sections: an introductory section to allow participants to warm up, covering general experiences with the app; a second section exploring users’ experiences with Quit Coach; and a final section exploring situations in which participants engaged with Quit Coach (Multimedia Appendix 1). Prompts were used to encourage participants to elaborate on their impressions and experiences. The topic guide was pilot-tested by KS on 2 graduate colleagues from the MSc program in Behavior Change and adapted following their feedback. The topic guide was further adapted following the interviews conducted in 2020 to facilitate elaboration by adding specific probes in addition to a new question (“To what extent would you say you formed a relationship of sorts with the chatbot? How was this?”). As participants mentioned their relationship with the chatbot, we considered it useful (and consistent with the flexible, semistructured style of interviewing) to specifically prompt subsequent participants about this. The adapted topic guide was piloted by AA on a graduate colleague from the MSc program in Behavior Change and a current smoker from AA’s network and updated according to their feedback. Interviews remained flexible, facilitated by the semistructured style of questioning.

**Procedure**

Upon expressing interest, participants were asked to read the participant information sheet, provide informed consent, and complete the eligibility questionnaire via Qualtrics (Qualtrics...
International Inc). If eligible, participants recruited in 2020 were contacted by the researchers to arrange the interview. If eligible, participants recruited in 2021 were contacted to confirm study acceptance and provided with information on how to download the pro version of the Smoke Free app (using a free access code). Participants were asked to select a quit date and nominate a day that was 2 weeks from their quit date to complete the interview.

Owing to the COVID-19 pandemic, in-person interviews were not possible. Therefore, interviews were conducted via the web by KS or AA (graduate students enrolled in an MSc program in Behavior Change) via Microsoft Teams. Besides the participant and the researcher, no one else was present during the interviews. KS had limited experience in conducting qualitative interviews before this study; AA had extensive experience from working for a consultancy firm. Before conducting the interviews, OP (PhD in Health Psychology, extensive experience in conducting qualitative interviews through previous academic work) provided training to KS and AA. Interviews were audio recorded and lasted between 30 and 45 minutes. Following completion, the participants were thanked, verbally debriefed, and presented with an incentive.

The Smoke Free App and Quit Coach

Smoke Free [35] is an evidence-informed app with a large user base (approximately 4000 downloads per day). The app contains behavior change techniques that are expected from theory and evidence from other settings to aid smoking cessation [17,36]. Refer to the study by Jackson et al [37] for a summary of the behavior change techniques included in the Smoke Free app, coded against a 44-item taxonomy of techniques used in individual behavioral support for smoking cessation [36].

The pro (ie, paid) version of Smoke Free contains a text-based, quick-response chatbot called Quit Coach (Figure 1). During the first 2 weeks of a user’s quit attempt, Quit Coach initiates twice-daily check-ins with users regarding their cessation attempt through a push notification. Check-in frequency is programmed to reduce after 1 month and cease entirely after 90 days (when users are anticipated to have quit smoking). Users engage with Quit Coach via text messages, selecting from prewritten responses. There are a few exceptions to this format, such as when users complete certain exercises that require free-text input (eg, typing the mantra, “not another puff, no matter what”), to occasionally type what is influencing their craving, or for providing feedback on whether they found a piece of advice useful (refer to Multimedia Appendix 1 for additional screenshots of such interactions). A bespoke Node.js natural language processing framework (adapted from freely available, state-of-the-art source code by Smoke Free’s developers) is used to map free-text inputs onto their likely intent—this constitutes the only machine learning element of the chatbot. Users can also initiate check-ins themselves by opening Quit Coach to record a craving or ask for assistance via a get help now toolkit, which provides different options for directing the conversation with Quit Coach. The conversational options include craving management, relapse, difficult situations, and withdrawal. Quit Coach’s communications (typically in text form, but also through emojis and Graphics Interchange Formats [GIFs]) contain information about the health consequences of smoking, quitting tips, and motivational messages, simulating a text message conversation.
Data Analysis

Interviews conducted in 2020 and 2021 were combined to form a single data set. Analysis was performed through an inductive thematic approach following the methodology by Braun and Clarke [38,39]: (1) familiarizing with the data, (2) generating initial codes, (3) searching for themes, (4) reviewing the themes, (5) defining and naming the themes, and (6) producing the report. We focused on latent (rather than semantic) meanings [39]. Although the Model of Supportive Accountability was used to inform the interview topic guide, terminology from this model was used for coding only if deemed relevant, with alternative codes considered throughout the process.

Interviews were coded using Microsoft Word in batches of 4 to 5 by AA to facilitate an iterative and reflexive approach. Each transcript was read multiple times to familiarize with the data. Then, initial codes were generated. Coded transcripts were reread following initial code generation and discussed with OP, adding or refining codes as appropriate. Then, the coded extracts were examined and used to generate preliminary themes. Next, a second, independent coder (another student from the same MSc program in Behavior Change) helped to assess coding reliability. The second coder coded 2 interviews, 1 each from the 2020 and 2021 samples, which were selected using a random number generator. The second coder was instructed to inductively code each interview. The resulting codes were compared with those of the first coder conceptually, rather than for perfect word matching. Discrepancies were discussed and reconciled. Then, themes were reviewed, refined, named, and agreed upon through discussion among AA, OP, and JB. During coding, the possibility for differences between the 2020 and 2021 samples was considered. Theoretical saturation was judged to have been reached after 12 interviews.

External Validation

A subsample of 14% (2/14) of randomly selected participants were contacted and agreed to read the results and comment on the congruence of the themes and narrative generated by the researchers with their own experiences. Both participants agreed with the researchers’ interpretations.

Reflexivity

The interviewers (women, White ethnicity, nonsmokers, and unfamiliar with most participants before the interview) felt that a good rapport was built with all the participants. Some were more immediately verbose, whereas others took a little time to open up, but did so with prompting and encouragement. For the first few interviews conducted, the interviewers closely followed the topic guide; however, as salient conversation topics emerged, a more discursive style of questioning was adopted to explore salient topics in great depth. Before commencing the interviews, participants were told about the goals of the study and that the interviewers were not directly involved in the development of
the Smoke Free app; however, participants were unaware of the interviewers’ smoking status or theoretical assumptions regarding user engagement or smoking cessation.

**Ethics Approval**

Ethics approval for this study was obtained from University College London’s Research Ethics Committee (CEHP/2020/579). Participants provided written informed consent before participating in the study.

**Results**

**Participant Characteristics**

A total of 40 participants completed the screening survey and were eligible to participate in the study. Of the 40 participants, 26 (65%) participants did not complete an interview, as they later decided that it was not the right time to quit, became uncontactable, or failed to attend the interview. Table 1 shows a summary of the demographic and smoking characteristics of the 35% (14/40) included participants.

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>Recruitment year</th>
<th>Age (years)(^a)</th>
<th>Country</th>
<th>Gender</th>
<th>Cigarettes per day at baseline</th>
<th>Smoking status at the time of interview</th>
<th>Quit Coach use</th>
<th>Number of past-year quit attempts</th>
<th>Use of app-based support to stop smoking (name of the app)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>2020</td>
<td>30</td>
<td>United Kingdom</td>
<td>Female</td>
<td>N/A(^b)</td>
<td>I have stopped smoking completely in the last year</td>
<td>Moderate</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P2</td>
<td>2020</td>
<td>20</td>
<td>United Kingdom</td>
<td>Male</td>
<td>N/A</td>
<td>I have stopped smoking completely in the last year</td>
<td>A lot</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P3</td>
<td>2020</td>
<td>33</td>
<td>United Kingdom</td>
<td>Female</td>
<td>N/A</td>
<td>I have stopped smoking completely in the last year</td>
<td>Extreme</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P4</td>
<td>2020</td>
<td>39</td>
<td>United Kingdom</td>
<td>Female</td>
<td>N/A</td>
<td>I have stopped smoking completely in the last year</td>
<td>A lot</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P5</td>
<td>2020</td>
<td>44</td>
<td>United Kingdom</td>
<td>Female</td>
<td>N/A</td>
<td>I have stopped smoking completely in the last year</td>
<td>A lot</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P6</td>
<td>2021</td>
<td>25-34</td>
<td>United Kingdom</td>
<td>Male</td>
<td>7</td>
<td>Quit(^c)</td>
<td>N/A</td>
<td>1</td>
<td>No (N/A)</td>
</tr>
<tr>
<td>P7</td>
<td>2021</td>
<td>18-24</td>
<td>France</td>
<td>Female</td>
<td>3</td>
<td>Cut down(^c)</td>
<td>N/A</td>
<td>3</td>
<td>Yes (Qwit)</td>
</tr>
<tr>
<td>P8</td>
<td>2021</td>
<td>18-24</td>
<td>France</td>
<td>Female</td>
<td>10</td>
<td>Cut down(^c)</td>
<td>N/A</td>
<td>1</td>
<td>No (N/A)</td>
</tr>
<tr>
<td>P9</td>
<td>2021</td>
<td>25-34</td>
<td>France</td>
<td>Male</td>
<td>12</td>
<td>Cut down(^c)</td>
<td>N/A</td>
<td>3</td>
<td>Yes (Smoke Free)</td>
</tr>
<tr>
<td>P10</td>
<td>2021</td>
<td>18-24</td>
<td>United Kingdom</td>
<td>Female</td>
<td>5</td>
<td>Cut down(^c)</td>
<td>N/A</td>
<td>0</td>
<td>No (N/A)</td>
</tr>
<tr>
<td>P11</td>
<td>2021</td>
<td>18-24</td>
<td>Mauritius</td>
<td>Female</td>
<td>10</td>
<td>Cut down(^c)</td>
<td>N/A</td>
<td>1</td>
<td>No (N/A)</td>
</tr>
<tr>
<td>P12</td>
<td>2021</td>
<td>25-34</td>
<td>India</td>
<td>Male</td>
<td>6</td>
<td>Cut down(^c)</td>
<td>N/A</td>
<td>2</td>
<td>No (N/A)</td>
</tr>
<tr>
<td>P13</td>
<td>2021</td>
<td>18-24</td>
<td>United Kingdom</td>
<td>Male</td>
<td>10</td>
<td>Quit(^c)</td>
<td>N/A</td>
<td>3</td>
<td>No (N/A)</td>
</tr>
<tr>
<td>P14</td>
<td>2021</td>
<td>18-24</td>
<td>United Kingdom</td>
<td>Male</td>
<td>8</td>
<td>Quit(^c)</td>
<td>N/A</td>
<td>1</td>
<td>No (N/A)</td>
</tr>
</tbody>
</table>

\(^a\) For participants recruited in 2021, age was measured as a range.

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

\(^c\) Ascertained qualitatively during the interview.
Themes

Overview
A total of three high-order themes were developed to capture the participants’ experiences with Quit Coach and the potential mechanisms underpinning user engagement: (1) anthropomorphism of and accountability to Quit Coach, (2) Quit Coach’s interaction style and format, and (3) users’ perceived need for support. Refer to Multimedia Appendix 1 for additional quotations. Figure 2 presents a thematic map of the themes. Here, anthropomorphism of Quit Coach, which is influenced by its interaction style and format, and users’ perceived need for support leads to feelings of accountability and increased engagement. Continued engagement with Quit Coach reinforces this accountability through a bidirectional relationship. In addition to this indirect link, Quit Coach’s interaction style and format directly influences users’ engagement.

Figure 2. Thematic map. Arrows indicate the direction of relationships and the + and – signs indicate their valence. Blue boxes relate to theme 1, green boxes relate to theme 2, and red boxes relate to theme 3.

Anthropomorphism of and Accountability to Quit Coach
Many users ascribed human-like characteristics, thoughts, and behavior to Quit Coach—despite the awareness that it was fully automated—following an interaction experience (eg, colloquial language, GIFs, and emojis) that closely mimicked those with peers and family members via SMS text messages or WhatsApp. Users’ anthropomorphism of Quit Coach was evident from almost all users frequently referring to it as an embodied entity (ie, him, them, or someone) rather than an object (eg, it, Quit Coach, or chatbot):

> It felt like you were really connected to someone. [P5; 2020; quit smoking]

> Throughout this conversation I have referred to “someone” rather than “something” quite a few times, which I haven’t done on purpose. [P6; 2021; quit smoking]

Users compared Quit Coach’s motivational and monitoring support with that of close friends or family members. Consequently, users described feeling that Quit Coach cared about them, which helped foster a relational bond and promoted...
a feeling of accountability to Quit Coach to succeed in cessation. Feeling that Quit Coach was fulfilling a supportive, social role justified the daily monitoring (ie, push notifications), which was experienced by some participants as very frequent, boring, or annoying. This social presence promoted continued app engagement, as users were motivated to succeed and willing to engage and cooperate with Quit Coach:

It felt like a friend or family member seeing how I was...It cared about whether I was smoking or not...You feel like you’ve got someone who cares enough [that you] stop. A reason not to...[P10; 2021; cut down smoking]

Many users even described Quit Coach’s support as superior to human friends or family members, owing to its immediate availability, single purpose (ie, smoking cessation support), and nonjudgmental tone of voice (particularly when reporting a lapse) and the relationship being 1-way (ie, users do not have to reciprocate support or worry about being boring or bothering Quit Coach). Such responses contradict the users’ comparison of Quit Coach’s support with that provided by friends or family members, indicating that they may instead have experienced the support as similar to that provided by a therapist or coach (ie, a 1-way rather than 2-way relationship). However, users’ perceptions of Quit Coach’s support as superior to that of human friends or family members appeared to serve as an advantage rather than to negate the human-like experience. A minority of users anthropomorphized Quit Coach to a lesser extent, owing to previous experience with using chatbots and great awareness of their synthetic nature. Nonetheless, they still embraced the support offered by Quit Coach:

[Friends/family] don’t wanna talk about it every evening...To have the QC for that purpose [is helpful]...You’re not always gonna be lucky enough to have someone who’s just gonna be there to just listen to what you want to talk about on any given evening. [P6; 2021; quit smoking]

It’s cool that he wasn’t judgemental...He says “lots of people do smoke again when they start to stop but it doesn’t mean that it’s their loss and that they need to start all over again, it’s just the first step”...The point is that long term you are trying not to smoke. [P8; 2021; cut down smoking]

Many users perceived having access to Quit Coach’s thoughts and feelings, reporting that they worried that Quit Coach would feel angry, upset, or disappointed if they did not check in or reported a lapse and wanting to please it by checking in regularly. Consequently, many users felt positively accountable to Quit Coach for frequent engagement and abstinence. For example, many users reported feeling proud or particularly motivated to engage when they could report not smoking. In addition, most users reported feelings of worry or shame about Quit Coach’s anticipated reaction if they had lapsed. For a few users, this caused an unintended consequence; their feelings were so significant that they reported completely avoiding checking in or lying in their reports. However, for most users, this anticipated worry or shame appeared to be beneficial, representing a source of motivation to stay quit. A few participants who failed in their quit attempt over time reported that they started ignoring check-ins, possibly owing to negative avoidance as the prompts would remind them of their failure:

I wanted to find an opportunity to make it happy. [P3; 2020; quit smoking]

I didn’t want [to tell that [I smoked] to the robot every morning and every afternoon...I felt that I was accountable to it. [P11; 2021; cut down smoking]

Some self-contradiction was observed in how accountable users reported feeling to Quit Coach. At some time points during the interviews, users stated feeling accountable to Quit Coach, but at different time points, they stated feeling accountable to themselves, friends, or family members. This contradiction typically arose after being specifically prompted about accountability. Users may have retrospectively changed how accountable they felt to Quit Coach, because the interviewer brought to the fore that Quit Coach was not real, despite an anthropomorphic experience:

The accountability thing was definitely my relationship with [people] rather than my relationship with the app. [P6; 2021; quit smoking]

**Quit Coach’s Interaction Style and Format**

Quit Coach’s interaction style and format appeared to both directly and indirectly (ie, by giving rise to anthropomorphic experience of Quit Coach) influence users’ engagement. Quit Coach’s motivational and positive tone of voice encouraged many users to stay on track by reminding them of and praising them for their progress. GIFs and emojis were used alongside written text messages to create a positive mood and inject humor, thus enhancing the motivational and positive tone beyond that created through written text alone. Many users noted that the GIFs and emojis promoted a human-like perception of Quit Coach:

There was a certain level of wanting to go back and get those little GIFs or whatever...I was always glad to go and have a check-in. [P6; 2021; quit smoking]

Engagement was generally driven by Quit Coach prompting check-ins rather than by the users themselves. The prompts were perceived as useful, as users may otherwise have forgotten to check in. Most users reported that the daily check-in time requirement was acceptable; it did not take up much of their day. During working hours, some users felt that check-ins were sufficiently short to be manageable, whereas some users were very busy, preferring to complete check-ins before or after working hours. This was supported by the ability to set preferred check-in times. Where users reported check-ins being long, it typically referred to a subjective experience of long, which was linked to boredom and lack of interest, often driven by repetition, forced choice, or limited opportunities for free-text inputs. Many users reported that forced choice made engagement feel less burdensome (ie, easy and less time-consuming), which contributed to check-ins being “just the right amount of time” and was particularly welcome when users were craving cigarettes. In contrast, many users reported that forced-choice interactions quickly became boring as they could not express what they wanted more precisely (as they would with a human).
This reduced the interest in engagement with Quit Coach, with many users indicating a preference for typing free-text questions and responses:

[If you’re] distracted by wanting a cigarette [it’s] just easier if you’ve got options in front of you to just pick one. [P10; 2021; cut down smoking]

It was a bit...Samey. Sometimes I would just kind of click through it all and not have to react as much...You feel like you’re just going through the motions a little bit rather than actually thinking about it. [P6; 2021; quit smoking]

Most users mentioned that they would have liked tailoring of the chatbot interactions on 2 levels. First, although Quit Coach broadly aligned with most users’ cessation motivations, it did not discuss the specific personal motives that users had inputted elsewhere in the app. Second, many users wanted Quit Coach to remember more about what worked for them and modify the messages and advice accordingly. Although users did not report the lack of desired tailoring to be particularly detrimental to engagement, most users indicated that enhanced tailoring could have a positive impact. The ability to set check-in times according to personal preferences or anticipated times of need promoted engagement. Several users agreed that if Quit Coach could prompt them to engage before or during triggering situations, it would be very useful:

It wasn’t really a tailored fit for myself. It talked about infertility and [that’s] not something that bothers me. [P13; 2021; quit smoking]

A reminder of what I’ve done so far [that worked] would be helpful. [P9; 2021; cut down smoking]

[It] would’ve been great if...he could send me a notification at [or before a] time [of anticipating being triggered] saying “you’re gonna be OK” or “you’re gonna make it!” [P7; 2021; cut down smoking]

**Users’ Perceived Need for Support**

For all users, the perceived need for support from Quit Coach appeared to be related to the frequency and intensity of cravings. Engaging with Quit Coach provided a useful behavioral alternative to smoking or attentional distraction from cravings. This was particularly helpful when check-ins aligned with strong cravings. In moments where users were not smoking or thinking about smoking, the need for support and, consequently, engagement interest appeared to be low. For some users, Quit Coach triggered cravings by reminding them of smoking:

It was [a] great thing to keep my hands [busy] [and] just give me time to let the craving pass. [P2; 2020; quit smoking]

When it was going well and when I didn’t smoke I just didn’t even use the app because I was OK...I was like I’m doing well so what can the app give me right now? [P8; 2021; cut down smoking]

As cravings reduced, users’ perceived need for support decreased, leading to reduced engagement interest. However, several users recruited in 2020 (all of whom had successfully quit smoking) reported engaging with Quit Coach even after their cravings reduced or disappeared. For these users, the need for support shifted from primarily needing a distraction from cravings to maintaining motivation to stay quit and reinforce ex-smoker identity. Self-selection bias may explain this prolonged engagement; users recruited in 2020 were already using Quit Coach, had successfully quit smoking, and self-identified as heavy Quit Coach users:

[Usage] kind of dwindle[d] down to kind of extreme need, its more about the check-ins in the morning...The chatbot doesn’t have much use now it’s like ten weeks or something since I stopped. [P2; 2020; quit smoking]

For users recruited in 2021 who were unsuccessful in cessation, the motivation to engage with Quit Coach decreased after a period of failing. These users felt discouraged and despondent and did not want to be reminded of their failure. Interestingly, this was sometimes accompanied by a reversal of the anthropomorphic experience (eg, referring to Quit Coach as a robot):

[Not succeeding in quitting] made me resent – not resent, that’s a big word – but made me not want to tell that to the robot every morning and every afternoon. [P11; 2021; cut down smoking]

**Discussion**

**Principal Findings**

Using a qualitative approach, this study aimed to explore smokers’ experiences and engagement with a quick-response chatbot implemented within a popular smoking cessation app. Users’ experiences with the chatbot were largely positive. Anthropomorphism of the chatbot (ie, ascribing human-like characteristics, thoughts, and behavior to the chatbot) was enabled by its specific interaction style and format (eg, positive message tone and quick and easy-to-complete check-ins) and users’ perceived need for support, which appeared to give rise to feelings of accountability to the chatbot and increased engagement. Our results build on and extend previous qualitative findings pertaining to users’ experiences with companion chatbots [20,29] to a chatbot specifically designed to support smoking cessation.

A previous experimental study has shown that social responses to computers—that is, a direct consequence of anthropomorphism—are common and relatively easy to generate (eg, by providing the computer with human-like attributes, such as a language output) [40]. However, according to the uncanny valley hypothesis, chatbot designers have a fine line to tread [41]. Strong feelings of affinity are generated by great human-like qualities in a robot or computer program up to a point. Once it becomes very similar to or indistinguishable from a real human, people’s reactions can reverse because they may find them creepy or eerie [41]. Although the quick-response Quit Coach in this study was relatively simple (eg, it did not allow many free-text inputs from users) and made it clear to users that it is an automated bot, the presence of social cues (eg, its positive message tone and communication format similar to...
text messaging with GIFs and emojis) appeared to be sufficient for generating social responses from many users without entering the territory of the uncanny valley. This is positive, as it implies that simple (and relatively low-cost) chatbots generate feelings of affinity and that more complex chatbots that can more closely mimic human interactions may not be necessary.

However, users mentioned that they would have liked it if Quit Coach tailored its questions and responses to their unique situations and momentary needs. Therefore, future studies would benefit from exploring—for example, through user-centered design activities and experimental studies—additional design elements that can enhance Quit Coach’s similarity to humans, as this may further promote user engagement. In addition, some design elements appeared to detract from users’ anthropomorphism of and engagement with Quit Coach, such as its repetitive questions and responses and forced-choice interactions. For users who struggled to stay quit, the repetitiveness and inflexibility of Quit Coach appeared particularly salient, sometimes leading to a reversal of the anthropomorphic experience. Previous studies indicate that users’ perceived need for support and the target behavior itself (eg, progress toward smoking cessation) are important for continued engagement [9,10,15]. Similarly, high perceived need for support may be important for users to suspend disbelief and anthropomorphize conversational agents within the health care domain. The Three-Factor Theory of Anthropomorphism [14] predicts that people are more likely to anthropomorphize nonhuman agents or objects when anthropocentric knowledge is readily accessible and applicable (ie, when knowledge about how humans interact, think, and feel is judged as relevant for the interaction), when motivated to be effective social agents (ie, motivation to master one’s environment by increasing its predictability and controllability), and when lacking a sense of social connection to other humans (ie, feeling lonely or isolated) [42]. Future studies would benefit from building on and empirically testing such a theory of anthropomorphism within the human-computer interaction domain, with a view to improving the design of future conversational and relational agents for health and well-being.

Our findings also lend partial support to the Model of Supportive Accountability [22] in that users reported feeling accountable to checking in and updating the nonjudgmental and supportive Quit Coach. Finding a balance between nonjudgmental tone of voice and human-like social cues to generate feelings of affinity and accountability (as discussed previously) may be important for future behavior change chatbots. However, trust and competence (which are additional cornerstones of the Model of Supportive Accountability) were largely missing from users’ accounts. It is plausible that competence (eg, legitimacy of the information provided) and trust (eg, data security and confidentiality) were already assumed by users who had either voluntarily downloaded the Smoke Free app from a digital marketplace—likely selecting an app they trusted among the myriad of available apps [15]—or were asked to download it based on recommendation from university researchers. Alternatively, trust and competence may not be necessary conditions for supportive accountability to arise within human-chatbot relationships; this should be further explored in future studies.

**Strengths and Limitations**

This study was strengthened by recruiting both experienced and novice app users, having a second coder to help in validating the coding, using external validation to ensure that the researchers’ interpretations aligned with participants’ narratives, and achieving theoretical saturation. However, this study also had several limitations. First, self-selection bias may limit the applicability of the findings to other populations and settings. For example, the 36% (5/14) of participants recruited in 2020 had quit smoking successfully and were considered as heavy Quit Coach users, and most participants were young (ie, aged 18-44 years). Second, we did not record any additional support used by participants during their quit attempts (eg, pharmacological support), which may have influenced their perceived need for support. Third, for pragmatic purposes, we did not record participants’ actual engagement with Quit Coach, but instead relied on self-reports. Going forward, triangulation of qualitative and quantitative findings would be an important addition to the research literature. Fourth, the study was conducted during the COVID-19 pandemic—a time of significant change in people’s life and work conditions, including their smoking behavior [43]—which may also limit the applicability of our findings to other periods and contexts.

**Implications for Research and Practice**

Findings from this study have both theoretical and practical implications. First, our results indicate that the Model of Supportive Accountability [22] may usefully be extended from human to human-like support within digital interventions. However, future studies should further explore the specific conditions under which chatbot interactions lead to feelings of accountability and whether accountability is more easily generated within human-to-human (rather than human-to-bot) interactions. Second, our findings suggest that chatbots for smoking cessation may benefit from including more variation in conversations to prevent boredom and incorporating different levels of tailoring (including context-sensitive tailoring).

**Conclusions**

Anthropomorphism of a quick-response chatbot implemented within a popular smoking cessation app appeared to be enabled by its interaction style and format (eg, positive message tone and quick check-ins) and users’ perceived need for support, which may have given rise to feelings of accountability and increased engagement.
Acknowledgments

OP receives salary support from Cancer Research UK (C1417/A22962). AA, KS, JB, and OP are members of Shaping Public Health Policies to Reduce Inequalities and Harm—a UK Prevention Research Partnership Consortium (MR/S037519/1). The authors gratefully acknowledge the funding.

Data Availability

Participants did not provide consent for their anonymized transcripts to be shared openly. Therefore, the qualitative data underpinning the analyses are available to bona fide researchers upon reasonable request.

Conflicts of Interest

JB has received unrestricted funding to study smoking cessation outside the submitted study from Johnson & Johnson and Pfizer, who manufacture smoking cessation medications. JB and OP are unpaid members of the scientific advisory board for the Smoke Free app.

Multimedia Appendix 1

Interview topic guide, additional screenshots of Quit Coach, and additional participant quotations.

References


Abbreviations

GIF: Graphics Interchange Format
Smartphone Ownership, Smartphone Utilization, and Interest in Using Mental Health Apps to Address Substance Use Disorders: Literature Review and Cross-sectional Survey Study Across Two Sites

Michael Hsu¹, MD; Bianca Martin¹, BA; Saeed Ahmed², MD; John Torous³, MBI, MD; Joji Suzuki¹, MD

¹Department of Psychiatry, Brigham and Women's Hospital, Boston, MA, United States
²West Ridge Center, Rutland Regional Medical Center, Rutland, VT, United States
³Digital Psychiatry Division, Department of Psychiatry, Beth Israel Deaconess Medical Center, Boston, MA, United States

Corresponding Author:
Michael Hsu, MD
Department of Psychiatry
Brigham and Women's Hospital
75 Francis Street
Boston, MA, 02115
United States
Phone: 1 617 732 6753
Fax: 1 407 550 7505
Email: mhsu24@emory.edu

Abstract

Background: In recent years, there has been increasing interest in implementing digital technologies to diagnose, monitor, and intervene in substance use disorders. Smartphones are now a vehicle for facilitating telepsychiatry visits, measuring health metrics, and communicating with health care professionals. In light of the COVID-19 pandemic and the movement toward web-based and hybrid clinic visits and meetings, it has become especially salient to assess phone ownership among individuals with substance use disorders and their comfort in navigating phone functionality and using phones for mental health purposes.

Objective: The aims of this study were to summarize the current literature around smartphone ownership, smartphone utilization, and the acceptability of using smartphones for mental health purposes and assess these variables across two disparate substance use treatment sites.

Methods: We performed a focused literature review via a search of two academic databases (PubMed and Google Scholar) for publications since 2007 on the topics of smartphone ownership, smartphone utilization, and the acceptability of using mobile apps for mental health purposes among the substance use population. Additionally, we conducted a cross-sectional survey study that included 51 participants across two sites in New England—an inpatient detoxification unit that predominantly treats patients with alcohol use disorder and an outpatient methadone maintenance treatment clinic.

Results: Prior studies indicated that mobile phone ownership among the substance use population between 2013 and 2019 ranged from 83% to 94%, while smartphone ownership ranged from 57% to 94%. The results from our study across the two sites indicated 96% (49/51) mobile phone ownership and 92% (47/51) smartphone ownership among the substance use population. Although most (43/49, 88%) patients across both sites reported currently using apps on their phone, a minority (19/48, 40%) reported previously using any apps for mental health purposes. More than half of the participants reported feeling at least neutrally comfortable with a mental health app gathering information regarding appointment reminders (32/48, 67%), medication reminders (33/48, 69%), and symptom surveys (26/45, 58%). Most patients were concerned about privacy (34/51, 67%) and felt uncomfortable with an app gathering location (29/47, 62%) and social (27/47, 57%) information for health care purposes.

Conclusions: The majority of respondents reported owning a mobile phone (49/51, 96%) and smartphone (47/51, 92%), consistent with prior studies. Many respondents felt comfortable with mental health apps gathering most forms of personal information and with communicating with their clinician about their mental health. The differential results from the two sites, namely greater concerns about the cost of mental health apps among the methadone maintenance treatment cohort and less experience with downloading apps among the older inpatient detoxification cohort, may indicate that clinicians should tailor technological
interventions based on local demographics and practice sites and that there is likely not a one-size-fits-all digital psychiatry solution.

(JMIR Form Res 2022;6(7):e38684) doi:10.2196/38684

KEYWORDS

smartphone; mobile phone; addiction; substance use; phone ownership; health equity; digital psychiatry; digital phenotyping; phone applications; substance abuse; mHealth; phone utilization; mental health; mindfulness; digital mental health

Introduction

Handheld phones evolved at a lightning pace over the past 2 decades. Once devices that were solely used to text or call, smartphones now connect millions through social media, track health metrics, and have GPS-sensing capabilities. Smartphones also have a burgeoning role in telepsychiatry, which has been widely adopted during the COVID-19 pandemic. Tens of thousands of mental health apps are available through app stores. Digital phenotyping, which involves using passively and continuously collected sensory and user data to track movement, phone utilization, and communication, has potential apps for relapse prediction in schizophrenia [1], relapse prediction in bipolar disorder [2], and depressive episode detection [3].

Smartphone apps have expanded and can be used for the diagnosis, monitoring, and treatment of substance use disorders [4]. For instance, Reset-O is the first US Food and Drug Administration–approved prescription mobile app for the treatment of opioid use disorder with evidence of improving abstinence and treatment retention [5]. Additionally, a smartphone app may be able to detect an opioid overdose by using a short-range active sonar [6]. Further, A-CHESS (Addiction-Comprehensive Health Enhancement Support System) is a mobile app suite that buttresses alcohol recovery through features such as delivering alerts to patients when they approach known triggers, such as a bar or liquor store, via GPS [7].

With smartphones playing an increasingly integral role in the delivery of substance use care, addressing equity and understanding the adoption and utilization of smartphones and the acceptability of related technologies have become ever more pressing. We performed a nonsystematic review of literature around smartphone ownership and utilization among individuals with substance use disorders, and we present original data from our 2-site, cross-sectional survey study assessing smartphone ownership, smartphone utilization, and the acceptability of using mental health apps. We hypothesize that a vast majority of individuals own smartphones and would be open to using smartphone apps to address substance use. The aim of this study is to inform clinicians who plan to develop or implement digital interventions for patients with substance use disorders on how to anticipate and optimize adoption and engagement.

Methods

Literature Review

The purpose of the literature review was to identify articles that assessed smartphone or phone ownership and/or utilization to answer the following questions: what proportion of patients with substance use disorders own mobile phones or smartphones, how did they utilize their phones, and how open were they to mobile health (mHealth) intervention through smartphones?

A nonsystematic search was conducted in January 2022 within PubMed and Google Scholar, using the key terms smartphone ownership, phone ownership, access to a smartphone, access to a smartphone, substance use disorder, addiction, and substance abuse, by two independent researchers. The selected publication types included primary survey or questionnaire studies, systematic or nonsystematic reviews, and secondary research studies based on prior survey or census data published in or after 2007 (the year that the Apple iPhone was introduced). Other inclusion criteria included articles written in the English language and participants with substance use disorders. The exclusion criteria included clinical trials of specific digital interventions, such as a smartphone app, as the purpose of this study was to understand general smartphone ownership and use. Other exclusion criteria included papers written in foreign languages and studies on participants with behavioral addictions, such as internet gaming disorder, internet addiction, or gambling disorder, as opposed to substance use disorders. A total of 8 abstracts were identified by using these criteria. Study purposes, study designs, descriptions of methods, sample sizes, sample demographics, smartphone or phone ownership rates, and smartphone utilization and acceptability data were extracted.

Cross-sectional, 2-Site Survey Study

We developed a survey that closely mirrored the one developed by Torous et al [8] (Textbox 1). The 5- to 10-minute survey was completed via paper and pencil and included sections on demographics, phone ownership, phone utilization, and the acceptability of using phones for mental health care purposes. Surveys were administered between June 2021 and February 2022 at two sites—a level 4 inpatient detoxification unit at a community hospital in Massachusetts (Brigham and Women’s Faulkner Hospital [BWFH]) that predominantly treats individuals with alcohol use disorder and an outpatient methadone clinic in Rutland, Vermont (West Ridge Clinic). Participants were offered a US $5 Dunkin’ Donuts gift card for completing this study. There was a total of 51 participants from both sites—22 from the inpatient detoxification unit and 29 from the outpatient methadone clinic. Descriptive statistics were used to summarize the data; all analyses were performed by using Stata Statistical Software: Release 17 (StataCorp LLC). Demographic and mobile phone use variables for those at the detoxification unit and those at the outpatient methadone clinic were compared by using a chi-square test for categorical variables and a 2-tailed Student t test for continuous variables.

https://formative.jmir.org/2022/7/e38684

JMIR Form Res 2022 | vol. 6 | iss. 7 | e38684 | p.112

(page number not for citation purposes)
### Questions

1. Do you currently own any type of phone?
   - If no, for what reason do you not have a phone?

2. Is your phone a smartphone (eg, iPhone, Android, etc)?

3. What is the name of your smartphone (eg, Samsung Galaxy)?

4. How comfortable are you sending text messages on your phone?

5. What type of payment plan do you use for text messages?

6. Do you download apps onto your phone?

7. Have you ever downloaded an app for your mental health?

8. Do you currently use any apps for your phone?

9. How comfortable or uncomfortable would you feel about a mental health app gathering and/or sending the following data from your smartphone to your clinician in the context of your care?
   - Appointment reminders
   - Medication reminders
   - Symptom surveys (eg, survey questions about your mood or thoughts throughout the day)
   - Your location (phone GPS sensor)
   - Your social information (call and text logs without any phone numbers or context of messages; eg, how many people you called and for how long)
   - Coaching for healthy living (eg, exercise, sleep, and diet)
   - Mindfulness or therapy exercises
   - Communicating with my clinician about my mental health

10. Select up to 3 top concerns you may have about mental health apps or apps for substance use disorders:
   - Privacy
   - Accuracy of recommendations from app
   - Hard to use
   - Sharing information with clinician
   - Cost
   - Time
   - Hard to set up

11. Select up to 3 top benefits you may see in mental health apps or apps for substance use disorders:
   - Privacy
   - Accuracy of recommendations from app
   - Easy to use
   - Sharing information with clinician
   - Cost
   - Time
   - Easy to set up

### Ethical Considerations

This study was approved and monitored by the Mass General Brigham Institutional Review Board (approval number 2020P001656) and Rutland Regional Hospital Institutional Review Board (approval number 2020P001656/RRMC24). This study was identified as human subjects research and was classified as exempt by the Mass General Brigham Institutional Review Board, given the minimal risk to subjects and the use of a survey tool with no identifiable information obtained. This study adheres to the ethical guidelines set forth by the Mass General Brigham Institutional Review Board.
General Brigham Human Research Protection Program [9]. Participants were consented prior to the survey study and were given the choice to opt out of specific questions or this study entirely at any time. No personal health or identifiable information was recorded.

Results

Participant Characteristics

For the inpatient detoxification cohort, the study team approached 35 participants, of whom 25 consented to answering the survey and 22 completed it (response rate: 22/35, 63%). Further, 2 participants were discharged from detoxification prior to completion, and 1 participant dropped out. The BWFH inpatient detoxification sample skewed toward male (12/22, 55%) and White (20/22, 91%) patients, and the West Ridge Clinic sample skewed toward female (18/29, 62%) and White (23/29, 79%) patients (Table 1). Notably, the West Ridge Clinic had a relatively higher proportion of individuals experiencing homelessness (10/28, 36%) compared to that of the inpatient detoxification clinic (1/22, 5%), though this did not reach statistical significance (P=.10). There were no additional differences in demographics between the two clinic sites except for the fact that the mean age at the methadone maintenance treatment (MMT) clinic was significantly lower (48.71 vs 36.04 years; P<.001).
<table>
<thead>
<tr>
<th>Variable</th>
<th>All participants (N=51)</th>
<th>BWFH\textsuperscript{a} inpatient detoxification clinic (n=22)</th>
<th>West Ridge Clinic (n=29)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>41.47 (13.0)</td>
<td>48.71 (11.8)</td>
<td>36.04 (11.2)</td>
<td>.001</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.24</td>
</tr>
<tr>
<td>Male</td>
<td>23 (45)</td>
<td>12 (55)</td>
<td>11 (38)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>28 (55)</td>
<td>10 (45)</td>
<td>18 (62)</td>
<td></td>
</tr>
<tr>
<td>Race/ethnicity, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.32</td>
</tr>
<tr>
<td>Black or African American</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>43 (84)</td>
<td>20 (91)</td>
<td>23 (79)</td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latinx</td>
<td>3 (6)</td>
<td>2 (9)</td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>3 (6)</td>
<td>0 (0)</td>
<td>3 (10)</td>
<td></td>
</tr>
<tr>
<td>Alaska Native</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td>Education (highest level completed), n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.20</td>
</tr>
<tr>
<td>Completed high school or General Educational Development</td>
<td>13 (25)</td>
<td>5 (23)</td>
<td>8 (28)</td>
<td></td>
</tr>
<tr>
<td>Some high school</td>
<td>7 (14)</td>
<td>1 (4)</td>
<td>6 (21)</td>
<td></td>
</tr>
<tr>
<td>Completed college or associate degree</td>
<td>6 (12)</td>
<td>3 (14)</td>
<td>3 (10)</td>
<td></td>
</tr>
<tr>
<td>Some college or associate degree</td>
<td>17 (33)</td>
<td>7 (32)</td>
<td>10 (34)</td>
<td></td>
</tr>
<tr>
<td>Graduate school</td>
<td>8 (16)</td>
<td>6 (27)</td>
<td>2 (7)</td>
<td></td>
</tr>
<tr>
<td>Work, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.18</td>
</tr>
<tr>
<td>Full-time employment</td>
<td>12 (23)</td>
<td>4 (18)</td>
<td>8 (28)</td>
<td></td>
</tr>
<tr>
<td>Part-time employment</td>
<td>5 (10)</td>
<td>1 (5)</td>
<td>4 (14)</td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>18 (35)</td>
<td>7 (32)</td>
<td>11 (38)</td>
<td></td>
</tr>
<tr>
<td>SSD\textsuperscript{b} or SSI\textsuperscript{c}</td>
<td>9 (18)</td>
<td>4 (18)</td>
<td>5 (17)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>3 (6)</td>
<td>2 (9)</td>
<td>1 (3)</td>
<td></td>
</tr>
<tr>
<td>Retired</td>
<td>4 (8)</td>
<td>4 (18)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Residence (n=50)\textsuperscript{d}, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.10</td>
</tr>
<tr>
<td>Own or rent apartment</td>
<td>18 (36)</td>
<td>10 (45)</td>
<td>8 (29)</td>
<td></td>
</tr>
<tr>
<td>Family or friends</td>
<td>10 (20)</td>
<td>6 (27)</td>
<td>4 (14)</td>
<td></td>
</tr>
<tr>
<td>Single room occupancy</td>
<td>8 (16)</td>
<td>4 (18)</td>
<td>4 (14)</td>
<td></td>
</tr>
<tr>
<td>Halfway house</td>
<td>3 (6)</td>
<td>1 (5)</td>
<td>2 (7)</td>
<td></td>
</tr>
<tr>
<td>Homeless</td>
<td>11 (22)</td>
<td>1 (5)</td>
<td>10 (36)</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}BWFH: Brigham and Women’s Faulkner Hospital.

\textsuperscript{b}SSD: Social Security Disability.

\textsuperscript{c}SSI: Supplemental Security Income.

\textsuperscript{d}One participant did not answer the question regarding place of residence.

### Phone Ownership in the Substance Use Population

Prior studies indicated that mobile phone ownership among the substance use population from 2013 to 2019 ranged from 83% to 94%, while smartphone ownership ranged from 57% to 94% [10-17] (Table 2). Individuals of the baby boomer generation (aged >52 years) may be less likely to own a mobile phone with app capability when compared to Generation X (age: range 36 to 51 years; P=.001; odds ratio 3.52, 95% CI 1.65-7.52) and millennials (age: range 18 to 35; P<.001; odds ratio 4.53, 95% CI 2.19-9.35) [14].

In our study, a large proportion of respondents reported owning a mobile phone (49/51, 96%) and smartphone (47/51, 92%). All patients (22/22, 100%) at the inpatient detoxification clinic reported owning a mobile phone, while 27 of 29 patients (93%) reported owning a mobile phone at the outpatient methadone clinic. Of the 49 respondents who owned mobile phones across both sites, 47 (96%) categorized them as smartphones; 2 participants opted out of this question.

https://formative.jmir.org/2022/7/e38684

JMIR Form Res 2022 | vol. 6 | iss. 7 | e38684 | p.115
(page number not for citation purposes)
Table 2. Mobile phone and smartphone ownership among individuals with substance use disorders across studies.

<table>
<thead>
<tr>
<th>Authors, year</th>
<th>Patient population</th>
<th>Sample size, N</th>
<th>Mobile phone ownership, %</th>
<th>Smartphone ownership, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>McClure et al, 2013 [16]</td>
<td>Adult patients who were undergoing substance abuse treatment and were enrolled at 8 drug-free psychosocial or opioid-replacement therapy clinics in Baltimore</td>
<td>266</td>
<td>91</td>
<td>N/A^2</td>
</tr>
<tr>
<td>Dahne and Lejuez, 2015 [12]</td>
<td>Adult patients admitted to a residential substance use treatment center in Washington, District of Columbia</td>
<td>251</td>
<td>86.9</td>
<td>68.5</td>
</tr>
<tr>
<td>Milward et al, 2015 [13]</td>
<td>Patients enrolled in 4 UK community drug treatment services (74% were undergoing treatment for heroin addiction)</td>
<td>398</td>
<td>83</td>
<td>57</td>
</tr>
<tr>
<td>Ashford et al, 2018 [14]</td>
<td>Adult patients in 4 intensive outpatient substance use disorder treatment facilities in Philadelphia</td>
<td>259</td>
<td>93.8</td>
<td>64.1</td>
</tr>
<tr>
<td>Curtis et al, 2019 [10]</td>
<td>Adolescents (aged 13-17 years) and emerging adults (aged 18-35 years) engaged in outpatient substance use treatment in the Southwest and Northeast regions of the United States</td>
<td>164</td>
<td>92.2</td>
<td>80.9</td>
</tr>
<tr>
<td>Tofghi et al, 2019 [15]</td>
<td>Adult patients enrolled in an inpatient detoxification program at a safety-net tertiary referral center in New York City</td>
<td>206</td>
<td>86</td>
<td>66</td>
</tr>
</tbody>
</table>

^2N/A: not applicable.

**Phone Utilization in the Substance Use Population**

Based on our review of the literature, 79% to 96% of individuals with substance use disorders have phones with text messaging capabilities [12,15,16], with one study published in 2015 suggesting that 55% of mobile phone owners use their phones to text daily [13]. Between 61% to 68% of individuals use their mobile phones to access the internet [10,12,15], and 61.3% of adult smartphone owners use mobile apps on their phones [12]. One study found that of individuals who accessed the internet, 80% accessed it primarily through their mobile phones, with no generational differences in terms of using phones to access the internet versus other means [14]. Masson et al [11] concluded that 40% (n=70) of their participants used their mobile devices as a reminder to take medications, while 8% (n=13) of smartphone users utilized a medication reminder app. McClure et al [16] found that 44% of their adult participants reported being called by substance use clinic staff, while 0.4% reported being texted by them.

Patients were also asked about their mobile phone use patterns in our study (Table 3). Nearly all patients across both sites (45/49, 92%) reported feeling extremely, very, or somewhat comfortable with sending text messages. Of the 47 patients who reported their type of text message payment plan, most (42/47, 89%) reported paying a flat fee for unlimited text messages. The remaining 5 patients reported either paying a flat fee for a limited text message plan (2/47, 4%) or using a pay-per-text plan (3/47, 6%). The majority of respondents (43/49, 88%) reported having downloaded apps, although this differed significantly across sites (West Ridge Clinic patients: 26/27, 96%; inpatient detoxification clinic patients: 17/22, 77%; P=.04). Although most patients (43/49, 88%) across both sites reported currently using apps on their phone, only a minority (19/48, 40%) reported having previously used any app for their mental health; this did not differ significantly across sites (P=.21).
<table>
<thead>
<tr>
<th>Variable</th>
<th>All participants (N=51), n (%)</th>
<th>BWFH(^a) inpatient detoxification clinic (n=22), n (%)</th>
<th>West Ridge Clinic (n=29), n (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mobile phone ownership</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smartphone (n=47)(^b)</td>
<td>49 (96)</td>
<td>22 (100)</td>
<td>27 (93)</td>
<td>.21</td>
</tr>
<tr>
<td></td>
<td>47 (100)</td>
<td>20 (100)</td>
<td>27 (100)</td>
<td></td>
</tr>
<tr>
<td><strong>Sending text messages (n=49)(^c)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extremely, very, or somewhat comfortable</td>
<td>45 (92)</td>
<td>19 (86)</td>
<td>26 (96)</td>
<td>.21</td>
</tr>
<tr>
<td><strong>Text message payment plan (n=47)(^d)</strong></td>
<td></td>
<td></td>
<td></td>
<td>.30</td>
</tr>
<tr>
<td>Flat fee for unlimited text messages</td>
<td>42 (89)</td>
<td>19 (95)</td>
<td>23 (85)</td>
<td></td>
</tr>
<tr>
<td>Flat fee for limited text messages</td>
<td>2 (4)</td>
<td>1 (5)</td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td>Pay-per-text plan</td>
<td>3 (7)</td>
<td>0 (0)</td>
<td>3 (11)</td>
<td></td>
</tr>
<tr>
<td><strong>Downloads apps onto phone (n=49)(^e)</strong></td>
<td></td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Has downloaded app for mental health</td>
<td>19 (40)</td>
<td>10 (48)</td>
<td>9 (33)</td>
<td>.32</td>
</tr>
<tr>
<td>Currently uses any apps on phone</td>
<td>43 (88)</td>
<td>18 (82)</td>
<td>25 (93)</td>
<td>.25</td>
</tr>
<tr>
<td><strong>Very, somewhat, or neutrally comfortable with mental health app gathering information (n=48)(^f)</strong></td>
<td>32 (67)</td>
<td>11 (52)</td>
<td>21 (78)</td>
<td>.06</td>
</tr>
<tr>
<td>Appointment reminders</td>
<td>33 (69)</td>
<td>12 (57)</td>
<td>21 (78)</td>
<td>.13</td>
</tr>
<tr>
<td>Medication reminders</td>
<td>26 (58)</td>
<td>11 (52)</td>
<td>15 (63)</td>
<td>.49</td>
</tr>
<tr>
<td>Location</td>
<td>18 (38)</td>
<td>7 (33)</td>
<td>11 (42)</td>
<td>.53</td>
</tr>
<tr>
<td>Social information</td>
<td>20 (43)</td>
<td>8 (38)</td>
<td>12 (46)</td>
<td>.58</td>
</tr>
<tr>
<td>Coaching for healthy living</td>
<td>27 (56)</td>
<td>11 (52)</td>
<td>16 (59)</td>
<td>.63</td>
</tr>
<tr>
<td>Mindfulness or therapy exercises</td>
<td>31 (65)</td>
<td>12 (57)</td>
<td>19 (70)</td>
<td>.34</td>
</tr>
<tr>
<td>Communicating with clinician about mental health</td>
<td>30 (63)</td>
<td>11 (52)</td>
<td>19 (70)</td>
<td>.20</td>
</tr>
<tr>
<td><strong>Perceived concerns about mental health apps</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Privacy</td>
<td>34 (67)</td>
<td>14 (64)</td>
<td>20 (69)</td>
<td>.69</td>
</tr>
<tr>
<td>Accuracy of recommendations</td>
<td>12 (24)</td>
<td>4 (18)</td>
<td>8 (28)</td>
<td>.43</td>
</tr>
<tr>
<td>Hard to use</td>
<td>9 (18)</td>
<td>5 (23)</td>
<td>4 (14)</td>
<td>.41</td>
</tr>
<tr>
<td>Sharing information with clinician</td>
<td>14 (28)</td>
<td>7 (32)</td>
<td>7 (24)</td>
<td>.54</td>
</tr>
<tr>
<td>Cost</td>
<td>15 (29)</td>
<td>2 (9)</td>
<td>13 (45)</td>
<td>.006</td>
</tr>
<tr>
<td>Time</td>
<td>16 (31)</td>
<td>5 (23)</td>
<td>11 (38)</td>
<td>.25</td>
</tr>
<tr>
<td>Hard to set up</td>
<td>11 (22)</td>
<td>7 (32)</td>
<td>4 (14)</td>
<td>.12</td>
</tr>
<tr>
<td><strong>Perceived benefits of mental health apps</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Privacy</td>
<td>11 (22)</td>
<td>4 (18)</td>
<td>7 (24)</td>
<td>.61</td>
</tr>
<tr>
<td>Accuracy of recommendations</td>
<td>14 (28)</td>
<td>3 (14)</td>
<td>11 (38)</td>
<td>.05</td>
</tr>
<tr>
<td>Easy to use</td>
<td>22 (43)</td>
<td>10 (46)</td>
<td>12 (41)</td>
<td>.77</td>
</tr>
<tr>
<td>Sharing information with clinianan</td>
<td>19 (37)</td>
<td>8 (36)</td>
<td>11 (38)</td>
<td>.91</td>
</tr>
<tr>
<td>Cost</td>
<td>8 (16)</td>
<td>2 (9)</td>
<td>6 (21)</td>
<td>.26</td>
</tr>
<tr>
<td>Time</td>
<td>20 (39)</td>
<td>8 (36)</td>
<td>12 (41)</td>
<td>.72</td>
</tr>
<tr>
<td>Easy to set up</td>
<td>18 (35)</td>
<td>4 (18)</td>
<td>14 (48)</td>
<td>.03</td>
</tr>
</tbody>
</table>

\(^a\)BWFH: Brigham and Women’s Faulkner Hospital.

\(^b\)Two participants who reported owning a mobile phone did not provide information about whether it was a smartphone.

\(^c\)Two participants who reported owning a mobile phone did not provide information about whether it was a smartphone.

\(^d\)Four participants did not answer questions regarding their comfort with sending text messages.

\(^e\)Two participants did not answer questions regarding their current text message payment plan.

\(^f\)Two participants did not answer questions regarding downloading apps onto their phones.
Acceptability of Using Smartphones for Mental Health Purposes

Based on our review of the literature, 70.4% of adult participants in one study stated that they would use a relapse prevention app [14], though a different study by Curtis et al [10] suggested that 36.9% of millennials and 45.3% of the Generation Z population thought that a mobile phone app would be helpful toward recovery. In another study, 46% of participants found it unacceptable to use geolocation in a smartphone app for health care purposes [13], but 86% of surveyed adults were willing to be contacted via mobile phone by their clinicians, with telephone calls (53%) being the most preferred method when compared to text messages (41%) and letters (41%) [13]. The most preferred frequency of contact was 1 to 2 messages weekly (46%) [13]. Further, 72% of adults surveyed by Ashford et al [14] in 2018 reported that it was acceptable to receive text messages for relapse prevention. Curtis et al [10] found that 28.8% of millennials and 45.3% of the Generation Z population thought that texting could be helpful toward recovery.

Over half of all participants at both sites were at least neutrally comfortable with a mental health app gathering information regarding appointment reminders (32/48, 67%), medication reminders (33/48, 69%), and symptom surveys (26/45, 58%). Most participants also found it acceptable to use mental health apps to engage in coaching for healthy living (27/48, 56%), mindfulness or therapy exercises (31/48, 65%), and communication with their clinician about their mental health (30/48, 62%). Notably, most of our sample expressed concerns about privacy (34/51, 67%) and reported being uncomfortable with an app gathering information about location (29/47, 62%) and social information (27/47, 57%) for health care purposes. The top three noted concerns about using mental health apps were privacy, cost, and time; patients at the outpatient methadone clinic were significantly more likely to perceive cost as a top-three concern (13/29, 45% vs 2/22, 9%; P=0.006). Overall, the top three perceived benefits of using mental health apps or apps for substance use disorders were ease of use, the ability to share information with clinicians, and time. Patients at the outpatient methadone clinic were also more likely to perceive ease of setup as a benefit of using such apps (14/29, 48% vs 4/22, 18%; P=0.03). The acceptability results are summarized in Table 3 and Figures 1-3.

Figure 1. Patients’ comfort with a mental health app gathering information on smartphone by clinic location. BWFH: Brigham and Women’s Faulkner Hospital.
Discussion

Principal Findings

The overall rate of mobile phone ownership was 96% (49/51), and the overall rate of smartphone ownership was 92% (47/51). The participants recruited at the community inpatient detoxification site were overall older (mean 48.71 vs mean 36.04 years; \( P < .001 \)), and a greater percentage were housed. Participants at the MMT clinic were more likely to report downloading apps when compared to the detoxification sample (26/27, 96% vs 17/22, 77%; \( P = .04 \)), which could be potentially explained by the younger mean age of the methadone cohort (36.04 vs 48.71 years; \( P < .001 \)) [18]. Most patients (43/49, 88%) reported regularly downloading and using apps on their phone, although only 40% (19/48) reported ever downloading an app specifically for mental health or substance use disorder purposes. A majority of participants across both clinic sites indicated feeling comfortable with mental health apps gathering most forms of personal information, specifically appointment reminders (32/48, 67%), medication reminders (33/48, 69%), symptom surveys (26/45, 58%), coaching for healthy living (27/48, 56%), mindfulness or therapy exercises (31/48, 65%), and communications with their clinician about their mental health (30/48, 62%). Most individuals were uncomfortable with a mental health app tracking location (29/47, 62%) or social information (ie, their call and text logs; 27/47, 57%). The differential views on cost as a barrier to using a mental health app across the two sites (methadone clinic: 13/29, 45%; inpatient detoxification clinic: 2/22, 9%; \( P = .006 \)) might reflect socioeconomic differences across the cohorts. As previously mentioned, the MMT sample had a relatively higher proportion of individuals experiencing homelessness (10/28, 36%) compared to that of the inpatient detoxification sample (1/22, 5%), though the comparative rates of homelessness across the
two samples did not reach statistical significance ($P=.10$). Most participants cited privacy (34/51, 67%) as a primary concern with using mental health apps and reported that they would not be comfortable with geolocation (29/47, 62%) or social information tracking (27/47, 57%). The participants recruited at the MMT site were significantly more likely to perceive the ease of setting up a mobile app as a perceived benefit when compared to those of the inpatient detoxification unit (14/29, 48% vs 4/22, 18%; $P=.03$), which may again speak to participants at the MMT clinic being more comfortable with using app functionalities.

Overall, our cross-sectional study suggests that individuals with substance use disorders are generally amenable to using a smartphone app for mental health monitoring or treatment purposes. Interestingly, while smartphone ownership was slightly lower among participants in the MMT site compared to that among the detoxification site participants, which is unsurprising given that the participants at the MMT site were of lower socioeconomic status, our data suggest that the individuals recruited at the MMT site had higher digital literacy, as reflected by their comfort with downloading apps and their perception that ease of use is a benefit of using a mental health app for substance use interventions. In conclusion, clinicians should consider patient demographics, digital literacy, and practice sites when implementing mHealth interventions for substance use disorders in an equitable fashion.

**Comparison to Prior Work**

To our knowledge, this study represents the first literature review of smartphone ownership, smartphone utilization, and the acceptability of using mHealth among individuals with substance use disorders and the first cross-sectional survey study to address this topic since the beginning of the COVID-19 pandemic. Smartphone and mobile phone ownership rates in our cross-sectional survey study were higher than those reported in all prior studies, likely reflecting the growing adoption of smartphones. Overall rates of downloading apps across both survey sites (43/49, 88%) were also higher than the 61% to 64% of participants who reported downloading mobile apps in a study by Dahne and Lejuez [12], which recruited patients from 2014 to 2015. The proportion of participants who felt uncomfortable with location being tracked (29/47, 62%) was slightly higher than that reported by Milward et al [13]. Comfort with utilizing specific functions of apps for substance use disorders, such as appointment reminders or social functions, and specific perceived benefits of using mobile apps for substance use disorders were not assessed in prior studies of individuals with substance use disorders.

**Strengths and Limitations**

This study has several limitations. First, we performed a brief, nonsystematic review, and it is likely that relevant papers may not have been included. We attempted to strengthen the robustness of this focused literature review by utilizing two independent reviewers, two separate search engines, and broad key words to capture and screen more abstracts. Future works should incorporate a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)-based systematic review to capture a broader range of studies. Second, the relatively small sample of our cross-sectional survey study precluded our ability to explore the impacts that race, socioeconomic status, age, and other factors have on smartphone ownership, utilization, and acceptability. Third, the predominantly White sample, especially the predominantly White inpatient detoxification cohort, limits the generalizability of this study. However, we recruited from two disparate clinical sites in two very different geographic locations to expand the diversity of recruited participants. Fourth, the degree of selection bias among our outpatient methadone clinic cohort is difficult to assess without ascertaining a survey response rate and consequently may impact the reliability of our results. However, we were able to obtain a survey response rate among individuals at our inpatient detoxification site. Those who turned down the survey at the inpatient detoxification site were asleep, were medically unwell, or were preoccupied at the time of survey distribution.

**Future Directions**

Future work in this area should include larger patient populations across various sites, which might include non–methadone outpatient substance use clinics. Further, in vivo, randomized controlled studies of promising mental health apps for substance use disorders are needed to establish clinical efficacy. Studies clarifying the effects of socioeconomic status, race, and other factors on digital literacy, smartphone utilization, smartphone ownership, and the acceptability of using apps for substance use interventions among individuals with substance use disorders are needed. Privacy and security concerns around mental health apps will need to be addressed, especially given that individuals with mental health and substance use disorders are particularly vulnerable.

**Acknowledgments**

We acknowledge Faith Stone from Vermont Blueprint for Health. This study was supported in part by the American Psychiatric Association/Substance Abuse and Mental Health Services Administration Minority Fellowship Program Grant, the Brigham Research Institute Microgrant, and a National Institutes of Health/National Institute on Drug Abuse grant (K23DA042326).

**Data Availability**

All data are represented in the tables and figures displayed in this publication. The data sets that were generated and/or analyzed during this study are available from the corresponding author on reasonable request.
References


https://formative.jmir.org/2022/7/e38684 JMIR Form Res 2022 | vol. 6 | iss. 7 | e38684 | p.121 (page number not for citation purposes)


Abbreviations

A-CHESS: Addiction-Comprehensive Health Enhancement Support System
BWFH: Brigham and Women’s Faulkner Hospital
mHealth: mobile health
MMT: methadone maintenance treatment
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

©Michael Hsu, Bianca Martin, Saeed Ahmed, John Torous, Joji Suzuki. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.07.2022. 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.
Feasibility of Mobile Health and Social Media–Based Interventions for Young Adults With Early Psychosis and Clinical Risk for Psychosis: Survey Study

Olivia H Franco¹, BS; Monica E Calkins¹, PhD; Salvatore Giorgi², MS; Lyle H Ungar², PhD; Raquel E Gur¹, MD, PhD; Christian G Kohler¹, MD; Sunny X Tang¹,³, MD

¹Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, United States
²Computer and Information Science, University of Pennsylvania, Philadelphia, PA, United States
³Feinstein Institutes for Medical Research, Northwell Health, Glen Oaks, NY, United States

Corresponding Author:
Sunny X Tang, MD
Feinstein Institutes for Medical Research
Northwell Health
75-59 263rd St
Glen Oaks, NY, 11004
United States
Phone: 1 631 786 3737
Email: stang3@northwell.edu

Abstract

Background: Digital technology, the internet, and social media are increasingly investigated as promising means for monitoring symptoms and delivering mental health treatment. These apps and interventions have demonstrated preliminary acceptability and feasibility, but previous reports suggest that access to technology may still be limited among individuals with psychotic disorders relative to the general population.

Objective: We evaluated and compared access to and use of technology and social media in young adults with psychotic disorders (PD), young adults with clinical risk for psychosis (CR), and psychosis-free youths (PF).

Methods: Participants were recruited through a coordinated specialty care clinic dedicated toward early psychosis as well as ongoing studies. We surveyed 21 PD, 23 CR, and 15 PF participants regarding access to technology and use of social media, specifically Facebook and Twitter. Statistical analyses were conducted in R. Categorical variables were compared among groups using Fisher exact test, continuous variables were compared using 1-way ANOVA, and multiple linear regressions were used to evaluate for covariates.

Results: Access to technology and social media were similar among PD, CR, and PF participants. Individuals with PD, but not CR, were less likely to post at a weekly or higher frequency compared to PF individuals. We found that decreased active social media posting was unique to psychotic disorders and did not occur with other psychiatric diagnoses or demographic variables. Additionally, variation in age, sex, and White versus non-White race did not affect posting frequency.

Conclusions: For young people with psychosis spectrum disorders, there appears to be no “technology gap” limiting the implementation of digital and mobile health interventions. Active posting to social media was reduced for individuals with psychosis, which may be related to negative symptoms or impairment in social functioning.

(JMIR Form Res 2022;6(7):e30230) doi:10.2196/30230

KEYWORDS

social media; psychosis; clinical high risk; technology; digital health
Introduction

Digital Health in Psychiatry
The increasing availability of technology has opened a new avenue for making health care more accessible to a broader population and for the development and implementation of digital interventions. In recent years, mobile and web-based health interventions have been used with great promise to assess and coordinate with mental health clients [1]. Social media and mobile health interventions present a unique opportunity to engage young people experiencing the emergence of a psychotic disorder (PD) or at clinical risk (CR) for psychosis, when there are clinically significant symptoms of psychosis, but threshold criteria are not yet met. The importance of early intervention and prevention in psychosis is a key factor in long-term success, and it has been shown that decreasing the duration of untreated psychosis increases long-term treatment outcome [2,3].

Mobile and social media interventions and assessment tools are relatively new in psychiatry, but initial studies show promising engagement, efficacy, and prognostic utility [4,5]. For example, a classification model based on social media language was able to identify early relapse signs for people experiencing first episode psychosis [6]. Relapse events were also predicted in another study by changes in smartphone-enabled social activities [7]. Among interventions, patients with PD benefitted from a combined treatment protocol involving monitoring via mobile health interventions and psychosocial services. Digital social networks have been integrated into mobile health apps to facilitate peer support and improve amotivation symptoms [8]. Other promising mobile health interventions include reminders for medication adherence and appointments, interventions, and case management via SMS text messages and apps on a smartphone. Mobile interventions can also vary from standardized communication to live interactions with clinicians; for example, the patient can have live sessions with a therapist via videoconferencing or SMS text messaging, or they can receive standardized communications like prerecorded videos, and tip sheets based on their input at the time of communication [9].

Access to Technology
Access to technology among the target population is critical for the feasibility of mobile and social media–based interventions. Globally, access to technology has increased dramatically across all age groups, but a “technology gap” may exist for individuals at risk for or diagnosed with psychotic disorders [2]. Among adult patients recruited from an inpatient and outpatient clinic, only 48% have access to the internet, and only 27% used social networking sites on a daily basis [10]. However, younger age may be associated with greater access to technology. In one study of young people in their first episode of psychosis, 100% used social media, and 90% of those reported daily use for an average of about 2 hours daily [5]. Another study completed in Spain showed that patients with first episode psychosis had similar interest in but decreased access to digital technologies including computers and smartphones [11]. If young people with psychosis commonly access digital technology and social media, it would provide an avenue for outreach, and could be a method of monitoring symptoms and side effects and providing treatment.

Objective
To our knowledge, no study has directly compared access to and use of technology and social media across the psychosis spectrum among CR, PD, and psychosis-free (PF) youths. Most studies also do not distinguish between passive access of social media and active posting. In this study, we surveyed CR and PD youths and PF comparison participants about their access to digital technology, the internet, and both access and posting to Facebook and Twitter. We hypothesized that the “technology gap” is decreased or absent among young adults with psychosis and clinical risk for psychosis, suggesting that social media and mobile health interventions would be feasible and implementable in this population.

Methods

Participants
For this study, three groups were evaluated: PD, CR, and PF. A total of 59 participants aged 18-32 years were included. PD participants were recruited from the University of Pennsylvania Psychosis Evaluation and Recovery Center, a coordinated specialty care clinic dedicated toward early psychosis. CR and PF participants were recruited among participants of other ongoing studies at the University of Pennsylvania and Children’s Hospital of Philadelphia Lifetime Brain Institute. We were interested in the main effect of psychosis on access to technology and use of social media. Because mental health disorders are common in the community for young adults and adolescents [12], our PF comparison group did not exclude people with nonpsychotic disorders.

All stable outpatients at the Psychosis Evaluation and Recovery Center aged 18-35 years as well as potential participants in Lifetime Brain Institute studies were approached and asked for their consent to participate in this survey as a screening procedure for larger studies. Participants all resided in the greater Philadelphia area, which includes both an urban and surrounding suburban environment. There was no remuneration for completing this survey. Several eligible individuals were not interested in research in general, but no one specifically refused to participate in this survey. This was a brief single timepoint assessment and no one withdrew prior to its completion.

Ethics Approval
All procedures were approved by the Institutional, Review Board at the University of Pennsylvania under protocol 831509, and the Children's Hospital of Philadelphia, protocol number 16-013305.

Study Design and Assessments
The survey was completed verbally with a research coordinator either in person or over the telephone. Individuals who agreed to participate were surveyed via a questionnaire regarding their access to technology and use of social media (Facebook and Twitter). Access to technology was defined generally as the ability to use the technology in a dependable manner, including...
personal ownership, shared devices, and public access points. Active posting was defined as commenting, posting status updates, or otherwise contributing original content, versus passive access of social media, which includes scrolling, viewing, or liking. We did not distinguish between accessing the internet via Wi-Fi or a data plan.

The PF and CR groups underwent semistructured interviewing with the Structured Interview for Prodromal Syndromes, which assessed threshold and subthreshold symptoms of psychosis. Determination for CR or PF status was made by consensus case conference. The PD group underwent consensus clinical diagnosis by psychiatrists and psychologists specializing in psychotic disorders and were determined to have one of the following Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) psychotic disorders: schizophrenia (n=13), schizoaffective disorder (n=4), bipolar I disorder with psychotic features (n=1), or unspecified psychotic disorder (n=3). Comorbid nonpsychotic disorders were evaluated based on DSM-5 criteria and were also present in all 3 groups, in proportion to expected population rates.

**Statistical Analysis**

Statistical analyses were conducted in R (version 3.5.2; R Foundation for Statistical Computing). Categorical variables were compared among groups using Fisher exact test, including the main outcomes of technology access and posting frequency versus group. Fisher exact test was also used to compare social media access and posting rates for other diagnoses as well as for sex and race effects. Age differences among the groups and between active versus nonactive users were compared using 1-way ANOVA. To account for group differences in age, sex, and race, we additionally performed logistic regressions predicting social media access and posting with group as the independent variable of interest, and covarying for the demographic variables. Significance was 2-tailed with $\alpha=0.05$.

**Results**

Table 1 displays details for 21 PD, 23 CR, and 15 PF participants. There was uniformly high access to mobile phones (96%-100%, $P>0.99$), smartphones (95%-100%, $P>0.99$), computers (85%-95%, $P=0.84$), and the internet (95%-100%, $P=0.61$). The majority of young adults accessed Facebook, but not Twitter. Social media access rates were similar for all 3 groups (62%-74% at least weekly, $P=0.73$). However, there was a significant main group effect for social media posting (5%-43% at least weekly, $P<0.009$). CR actively posted at a similar rate compared to PF individuals (CR 43% vs PF 27%, $P=0.31$), but PD actively posted at a significantly lower rate than the nonpsychotic groups (5%, $P<0.01$). When examining Facebook use alone, as this was the platform with the most users, we found consistent results: access was not affected by psychosis group ($P>0.99$), but there was a significant main effect for group on posting ($P=0.02$). When covarying for sex, age, and race, psychosis diagnosis remained a significant predictor of decreased active social media posting (standardized $\beta=-1.22$, $P=0.03$), but it was not a significant predictor for social media access (standardized $\beta=-0.29$, $P=0.40$).

Decreased active social media posting was unique to psychotic disorders and there was no group effect for diagnosis with attention-deficit/hyperactivity disorder ($P=0.67$), mood ($P=0.54$), or anxiety disorders ($P>0.99$) when comparing prevalence of weekly or greater use. Variation in age ($P=0.63$), sex ($P=0.13$), and White versus non-White race ($P=0.77$) did not affect posting frequency.
Table 1. Sample characteristics, access to technology, and social media use.

<table>
<thead>
<tr>
<th>Variable</th>
<th>PF&lt;sup&gt;a&lt;/sup&gt; (n=15)</th>
<th>CR&lt;sup&gt;b&lt;/sup&gt; (n=23)</th>
<th>PD&lt;sup&gt;c&lt;/sup&gt; (n=21)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>23.5 (4.2)</td>
<td>22.0 (2.9)</td>
<td>24.2 (3.4)</td>
<td>.10</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>11 (73)</td>
<td>12 (52)</td>
<td>5 (24)</td>
<td>.01</td>
</tr>
<tr>
<td>Male</td>
<td>4 (27)</td>
<td>11 (48)</td>
<td>16 (76)</td>
<td></td>
</tr>
<tr>
<td><strong>Race, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>4 (29)</td>
<td>11 (50)</td>
<td>14 (67)</td>
<td>.09</td>
</tr>
<tr>
<td>African American</td>
<td>7 (50)</td>
<td>9 (41)</td>
<td>6 (29)</td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td>1 (7)</td>
<td>2 (9)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>2 (14)</td>
<td>0 (0)</td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td><strong>Comorbidities, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention-deficit/hyperactivity disorder</td>
<td>2 (13)</td>
<td>3 (13)</td>
<td>2 (10)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>Anxiety disorder</td>
<td>2 (13)</td>
<td>15 (65)</td>
<td>5 (24)</td>
<td>.002</td>
</tr>
<tr>
<td>Mood disorder</td>
<td>3 (20)</td>
<td>12 (52)</td>
<td>7 (33)</td>
<td>.14</td>
</tr>
<tr>
<td><strong>Access to technology, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobile phone</td>
<td>15 (100)</td>
<td>22 (96)</td>
<td>21 (100)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>Smartphone</td>
<td>15 (100)</td>
<td>22 (96)</td>
<td>20 (95)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>Computer</td>
<td>13 (87)</td>
<td>21 (91)</td>
<td>20 (95)</td>
<td>.84</td>
</tr>
<tr>
<td>Internet</td>
<td>15 (100)</td>
<td>23 (100)</td>
<td>20 (95)</td>
<td>.61</td>
</tr>
<tr>
<td><strong>Social media use, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At least weekly access</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facebook</td>
<td>10 (67)</td>
<td>17 (74)</td>
<td>13 (62)</td>
<td>.73</td>
</tr>
<tr>
<td>Twitter</td>
<td>3 (20)</td>
<td>5 (22)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>At least weekly posting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facebook</td>
<td>4 (27)</td>
<td>10 (43)</td>
<td>1 (5)</td>
<td>.01</td>
</tr>
<tr>
<td>Twitter</td>
<td>2 (13)</td>
<td>3 (13)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Facebook (ever)</td>
<td>14 (93)</td>
<td>17 (74)</td>
<td>15 (71)</td>
<td>.27</td>
</tr>
<tr>
<td>Twitter (ever)</td>
<td>4 (27)</td>
<td>6 (26)</td>
<td>2 (10)</td>
<td>.30</td>
</tr>
</tbody>
</table>

<sup>a</sup>PF: psychosis-free.
<sup>b</sup>CR: clinical risk for psychosis.
<sup>c</sup>PD: psychotic disorder.

**Discussion**

**Principal Findings**

We found that youths with PD and CR who participate in our clinical care and research programs have similar access to technology and use social media to a similar degree compared with PF individuals. This supports the feasibility of mobile health and social media interventions in young CR and PD populations in terms of access to technology. Our results were consistent with those found in previous studies, which showed that the use of digital technology in the treatment of people with psychosis spectrum disorder and other chronic serious mental illnesses is a viable method of delivery of services. In our study, the absolute rate of mobile phone and smartphone ownership and internet use was higher than those reported by Young et al [13], likely because of the younger age of our participants. Most previous studies have compared access to technology between mental health clients and published normative data for the general population but have not directly compared across groups using a consistent standardized methodology. One study directly compared the use of (not access to) technology between patients with first episode psychosis and healthy control subjects; they found significant but small to moderate effects of decreased frequency of use for computers, tablets, smartphones, and smart televisions (but not game consoles) [11]. However, it is unclear whether demographic differences may have accounted for some of the disparity. No studies have included a psychosis clinical risk comparison group.
Another relevant consideration is comfort with technology among clinicians. A recent study by Camacho and Torous [14] surveyed 42 Coordinated Specialty Care clinics and found that health care providers were supportive of implementing technology in their care model for early psychosis; although 69% of surveyed staff were confident in their ability to provide technical assistance for others, 78% indicated that additional digital skills training would be beneficial.

We also found a psychosis-specific decrease in active social media posting. Social cognitive impairments and negative symptoms may contribute to this finding. This interpretation is supported by a study by Rehki et al [15] showing that an increase in severity of negative symptoms is associated with a lower likelihood of social media use. In that study, they suggested that social interactions via social media should be considered in the clinical evaluation of individuals with schizophrenia, as it is a prominent form of communication and fostering relationships. Although individuals with CR may also experience significant negative symptoms, our results suggest that these are not reflected in significant changes in social media posting and access rates. Although youths with PD post significantly less frequently than youths with CR and PF, individuals with PD nevertheless access social media at a similar rate, which provides a potential avenue for intervention. To our knowledge, no other study has distinguished between social media access and posting, so we are the first to report a decrease in social media posting for individuals with psychotic disorders while there was no difference in level of passive use of social media. This effect did not appear with mood disorders, attention-deficit/hyperactivity disorder, and anxiety, and therefore appears to be specific to psychosis.

Limitations of this study include sampling of one geographical area and limited sample size. Socioeconomic status may also influence access to technology but was not consistently measured in this sample. It is possible that patients with access to a specialized psychosis treatment center and those who volunteer for research participation may have higher access to technology than the general population of people with psychosis. Moreover, our survey only included two social media platforms, Facebook and Twitter. Future studies may consider including access and use of other social media platforms that have recently become more popular among youths, as well as questions about their willingness to share their digital access and to receive therapeutic interactions via a digital platform. Objective usage statistics and other metadata may also provide useful signals. Privacy risks and concerns may also be technical as well as subjective barriers to implementation of such interventions. These were not addressed with this study.

Conclusions

Overall, our findings encourage further development of mobile health and social media–based interventions and monitoring for young people with psychosis and youths at clinical risk for psychosis. There appears to be no significant “technology gap” for young people with psychotic disorders relative to young people without psychosis. Lower active engagement may reflect impairments in social cognition and functioning. Notably, access to technology does not mean that digital health interventions will be ultimately efficacious or implementable—however, the availability of technology to young people with psychosis does provide some basis for feasibility. Future studies are needed to directly evaluate the efficacy and usability of digital health and social media–based strategies for intervention and assessment, while considering additional potentially harmful effects related to privacy concerns and increased time spent on social media.

Acknowledgments

SXT received funding from the Brain and Behavioral Research Foundation NARSAD Young Investigator Grant and the ASCP Early Career Research Award. We thank the participants and their families for their collaboration. We also thank Lyndsay Schmidt, Zeeshan Huque, LaTonya McCurry, Victoria Pietruszka, Kosha Ruparel, Thomas Hohing, and Eehwa Ung from the University of Pennsylvania.

Conflicts of Interest

SXT is a consultant for Neurocrine Biosciences and North Shore Therapeutics, received funding from Winterlight Labs, and holds equity in North Shore Therapeutics. The other authors have no disclosures to report.

References


Abbreviations

CR: clinical risk for psychosis
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
PD: psychosis disorder
PF: psychosis-free
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.
Development of a Maternal and Child mHealth Intervention With Aboriginal and Torres Strait Islander Mothers: Co-design Approach

Sarah Jane Perkes1*, BSc, MPH; Belinda Huntriss2*, BEd; Noeline Skinner2*, ATSIHP; Bernise Leece2, ATSIHP; Rosie Dobson3, PhD; Joerg Mattes2, MBBS, PhD; Kerry Hall1, PhD; Billie Bonevski1, PhD

1Flinders Health and Medical Research Institute, College of Medicine and Public Health, Flinders University, Bedford Park, South Australia, Australia
2School of Medicine and Public Health, College of Health, Medicine and Wellbeing, University of Newcastle, Newcastle, Australia
3National Institute for Health Innovation, The University of Auckland, Auckland, New Zealand
4First Peoples Health Unit, Griffith University, Southport, Australia
*these authors contributed equally

Corresponding Author:
Sarah Jane Perkes, BSc, MPH
Flinders Health and Medical Research Institute
College of Medicine and Public Health
Flinders University
Sturt Rd
Bedford Park, South Australia, 5042
Australia
Phone: 61 451994827
Email: sarah.perkes@newcastle.edu.au

Abstract

Background: Despite their growing popularity, there are very few mobile health (mHealth) interventions for Aboriginal and Torres Strait Islander people that are culturally safe and evidence based. A co-design approach is considered a suitable methodology for developing health interventions with Aboriginal and Torres Strait Islander people.

Objective: The aim of this study was to co-design an mHealth intervention to improve health knowledge, health behaviors, and access to health services for women caring for young Aboriginal and Torres Strait Islander children.

Methods: Aboriginal researchers led engagement and recruitment with health services and participants in 3 Aboriginal and Torres Strait Islander communities in New South Wales, Australia. Focus groups and interviews were facilitated by researchers and an app developer to gather information on 3 predetermined themes: design characteristics, content modules, and features and functions. Findings from the co-design led to the development of an intervention prototype. Theories of health behavior change were used to underpin intervention components. Existing publicly available evidence-based information was used to develop content. Governance was provided by an Aboriginal advisory group.

Results: In total, 31 mothers and 11 health professionals participated in 8 co-design focus groups and 12 interviews from June 2019 to September 2019. The 6 design characteristics identified as important were credibility, Aboriginal and Torres Strait Islander designs and cultural safety, family centeredness, supportive, simple to use, and confidential. The content includes 6 modules for women’s health: Smoke-free families, Safe drinking, Feeling good, Women’s business, Eating, and Exercising. The content also includes 6 modules for children’s health: Breathing well; Sleeping; Milestones; Feeding and eating; Vaccinations and medicines; and Ears, eyes, and teeth. In addition, 6 technology features and functions were identified: content feed, social connection, reminders, rewards, communication with health professionals, and use of videos.

Conclusions: An mHealth intervention that included app, Facebook page, and SMS text messaging modalities was developed based on the co-design findings. The intervention incorporates health behavior change theory, evidence-based information, and the preferences of Aboriginal and Torres Strait Islander women and health professionals. A pilot study is now needed to assess the acceptability and feasibility of the intervention.

(JMIR Form Res 2022;6(7):e33541) doi:10.2196/33541

KEYWORDS
mHealth; co-design; Aboriginal and Torres Strait Islander; mother; baby; young children; mobile phone

https://formative.jmir.org/2022/7/e33541
Introduction

Background
The health and well-being of Aboriginal and Torres Strait Islander people have been significantly affected by dispossession, interruption of culture, and intergenerational trauma since the colonization of Australia [1]. The ongoing impact has resulted in an unequal opportunity for good health. The life expectancy of Aboriginal and Torres Strait Islander women is 8 years less than that of non-Indigenous Australian women [2]. In the 2018-2019 National Aboriginal and Torres Strait Islander Health Survey, the majority of women aged ≥15 years were not meeting guidelines for physical activity, vegetable intake, or fruit intake; 36% reported that they smoked tobacco daily; and 35% reported that they experienced very high levels of psychological distress [3]. Infant mortality continues to be unacceptably high for Aboriginal and Torres Strait Islander babies at 2.1 times the rate of non-Indigenous infants (6.3 and 3.1 per 1000 live births, respectively) [4]. Mothers and babies getting the best possible care and support for a good start to life is 1 of 12 health priorities of the National Aboriginal and Torres Strait Islander Health Plan 2013-2023 [1].

Aboriginal and Torres Strait Islander people make up 3.3% (798,400/24,193,939) of the Australian population [5] and include many distinct groups with their own language and culture. In total, 44% of Aboriginal and Torres Strait Islanders live in regional areas, 37% in cities, and 18% in remote or very remote areas [5]. Those living in regional and remote areas have less access to primary health care and overall poorer health [6]. Nationally, Aboriginal and Torres Strait Islander people have less access to the internet at home (75.3% compared with 85.8% of all Australians); there are significant differences based on location: 82.8% in cities, 73.2% in regional areas, 61.3% in remote areas, and 49.9% in very remote areas [7]. More than 1 in 3 (35%) Aboriginal and Torres Strait Islander people are mobile-only users compared with a national rate of 1 in 5 (19.9%); these figures are linked to socioeconomic factors [8]. Using only a mobile is likely to incur more costs for data, less capability, and less access to more sophisticated digital health information and tools [8]. It is of importance that mobile health (mHealth) interventions are developed with a goal to increase digital inclusion.

mHealth is the use of mobile technology to improve health. Functions include SMS text messaging, multimedia messaging service, voice, internet access, and software apps, which range in complexity. mHealth is used for a variety of purposes, including health education, health behavior change, sensors and point-of-care diagnostics, registries and vital-event tracking, and data collection [9]. mHealth is being used increasingly for health promotion because of its reach, with >7 billion mobile phone subscriptions globally [10]; the Be He@lthy, Be Mobile initiative by the World Health Organization has reached >3.5 million people [11]. There are limited recent national figures on smartphone ownership among Aboriginal and Torres Strait Islander people, although available data indicate that ownership is high: a survey with 400 Aboriginal and Torres Strait Islander people in 2014 reported that 70% of Aboriginal and Torres Strait Islander people owned a smartphone and 69% used Facebook compared with 66% and 40% respectively for non-Indigenous Australians [12]. The top reason for using a mobile phone in this group was to send SMS text messages [12].

Studies focused on Aboriginal and Torres Strait Islander people using SMS text messaging to improve health show high acceptability of the modality [13-15]. SMS text messaging has the advantage of being accessible on all mobile phones and not requiring access to a data service. There are few technical barriers to SMS text messaging and high acceptability of the modality among new mothers [16,17]. In a metareview (23 systematic reviews, 371 studies, and 79,665 participants) on the impact of mHealth on a range of outcomes, including clinical outcomes, adherence to treatment and care, health behavior change, disease management, and attendance rates, SMS text messaging was the most frequently examined function and reported to be the most successful overall [18]. SMS text messaging seems to be particularly effective at increasing smoking cessation rates (in adult smokers from mostly high-income countries) [19]. The evidence for SMS text messaging helping to improve nutrition and physical activity is not as strong; however, SMS text messaging used in conjunction with other mHealth functionality has shown significant positive effects for healthy eating [18].

Health apps continue to be popular, although the evidence suggests that apps have limited effectiveness on changing health behaviors [18,20-22]. Some studies have found that apps can be effective at changing behavior among some clinical groups [18], although overall there is limited evidence to date. Of the few trials focused on Indigenous populations, app use has been reported to be low [23,24]. A recent pilot randomized controlled trial of a smoking cessation app with 49 Aboriginal people in Australia reported low to moderate level of app use, and at 6-month follow-up, only 1 participant was abstinent [24]. The authors concluded that although there was broad acceptability for the app, mHealth interventions should be designed with functions that are commonly used, including social media platforms [24]. A co-designed mHealth app developed in New Zealand with Māori and Pacific Islander people was tested in a cluster randomized controlled trial in 2019 (n=1451) [23]. Adherence to health-related–behavior guidelines increased at 12 weeks in both groups, with no difference between the groups. Engagement with the app overall was low, although those who did engage with the app as it was designed saw greater benefit. The co-design approach was reported to have drawn a very positive response from the community, as was reflected in the high participation and follow-up rates [23].

Social media is a form of mHealth, with potential to support health. The Aboriginal and Torres Strait Islander health sector was an early adopter of social media networks to promote health [25,26]. Social media campaigns on COVID-19 by Aboriginal and Torres Strait Islander health organizations is a recent example [27]. A recent Cochrane review on behavioral interventions delivered through social media for health behavior change, health outcomes, and health equity (88 studies; n=871,378) reported varied effects; overall, social media was found to improve physical activity, weight loss, and general
well-being, and small to no effects were found for other outcomes [28]. No studies focusing on Aboriginal and Torres Strait Islander people were included in the review.

**Objectives**

In response to the limited mHealth interventions available for Aboriginal and Torres Strait Islander women and children, we aimed to co-design a prototype focused on the needs and ideas of Aboriginal and Torres Strait Islander mothers. Co-design is a partnership approach where end users are actively involved from conception to dissemination [29]. Using co-design methodologies is one of the guiding principles of the Aboriginal Health and Medical Research Council of New South Wales (NSW) Ethical Guidelines for conducting health research with Aboriginal people [30]. In this paper, we describe the co-design processes and findings, as well as provide a description of the mHealth prototype.

**Methods**

**Study Design**

In total, 8 focus groups and 12 interviews were conducted from June 2019 to September 2019. Surveys were used to collect demographics at the start of focus groups and interviews. An Aboriginal advisory group that included Aboriginal team members who were also members of the participating communities met quarterly to oversee design, implementation, analysis, and reporting. An expert mHealth research group was consulted for opinion on research and intervention design.

**Ethics Approval**

Human research ethics approval was received from the Aboriginal Health and Medical Research Council (1485/19) and the University of Newcastle (H-2019-00760).

**Co-design Framework**

A co-design framework for an mHealth intervention with Māori and Pacific communities in New Zealand [29] based on work by Bratteteig et al [31] was used to guide the methods used in this study. Co-design is a coherent methodology with a range of tools and techniques used to favor the preferences of end users [31]. The co-design methods used included focus group and interview discussions, card sorting, storyboarding, design activities, survey, guidance from expert groups, and an iterative design phase with the research team.

**Setting**

Focus groups and interviews were held at 3 regional NSW locations: Newcastle, Coffs Harbour, and Inverell. In total, 5 Aboriginal organizations (including 3 Aboriginal health services, an Aboriginal preschool, and an Aboriginal corporation) and 3 NSW Health sites participated. Venues for focus groups and interviews were decided in consultation with participants.

**Participants**

Women aged ≥16 years who were either mothers or primary carers of an Aboriginal or Torres Strait Islander child aged 0 to 5 years or were pregnant (≥20 weeks gestation), owned or regularly used a smartphone, and had accessed a participating service (Aboriginal health service or NSW Health service) were eligible to participate. Health professionals at participating services who worked with women or children were eligible.

**Procedures**

Convenience sampling was used to recruit participants. Aboriginal researchers (BH, NS, and BL) who worked within the participating communities used their personal networks. In addition, participants were asked if they would like to recommend a friend or family member to the study. Potential participants were screened for eligibility when they contacted the researcher on the telephone. The researcher explained the study and gained informed consent over the telephone initially and again in person before the start of the focus group or interview. Participants were reimbursed with a shopping voucher worth Aus $30 (US $21.6) for attending focus groups and interviews and provided with refreshments. Health professionals were recruited using a snowball methodology through the participating services. Health professionals were not reimbursed.

Mothers and health professionals participated in separate focus groups and interviews. Focus groups and interviews were cofacilitated by a combination of Aboriginal researchers (NS and BH), a PhD student (SJP), and an app developer. Interviews and focus groups were 20 to 90 minutes in length. The number of participants in focus groups ranged from 2 to 6. Focus groups and interviews were recorded and transcribed, and field notes were taken.

**Measures**

Different surveys and discussion guides were used with mothers and health professionals. Discussions and activities were used to identify (1) design characteristics, (2) content modules, and (3) features and functions.

**Mothers**

**Survey**

The survey comprised 16 items, including demographic, cultural, and socioeconomic items. The items were selected from a previous study [32], with all items having been tested with Aboriginal and Torres Strait Islander mothers previously.

**Discussion Guide**

In all focus groups and interviews with mothers, 3 main questions were asked. Follow-up questions were asked depending on responses. Additional questions about mobile phone use to inform features and functions were asked in focus groups cofacilitated by the app developer. The three main questions were as follows:

- How would an mHealth intervention designed for healthy living for Aboriginal and Torres Strait Islander people differ from other mHealth interventions?
- Are you more interested in mHealth for your own health or your child’s health? What topics and features interest you?
- What do you think stops or prevents some women from accessing health information and services for themselves and their children?
Activities
Card-sorting activities were used to identify current mobile phone use (functions used, frequency of use, and reasons for use). Storyboarding activity was used to elicit creative descriptions of the mHealth intervention using drawings and words on what the intervention should include. Design activity was used to gain feedback on potential designs.

Health Professionals
Survey
The survey comprised 5 items related to demographic and professional practice characteristics.

Discussion Guide
In all focus groups and interviews with health professionals, 3 main questions were asked. Additional follow-up questions were asked depending on the response. The three main discussion questions were as follows:
1. What do you think are the most important health and well-being topics to include for Aboriginal or Torres Strait Islander women, children, and family?
2. What are the barriers for Aboriginal or Torres Strait Islander families to having good health?
3. What types of mobile technology do you think could support Aboriginal or Torres Strait Islander women’s and children’s health?

Co-design Analysis
A generalized thematic analysis was completed. An Aboriginal researcher (BH) and a PhD student (SJP) independently coded themes. NVivo software (version 12.0; QSR International) was used to complete independent coding and comparison by the 2 coders. In total, 3 predetermined codes were used based on a similar co-design study [29]. These codes included (1) design characteristics, (2) content modules, and (3) features and functions. The coders met to agree on subcodes and definitions. Survey findings are presented using descriptive statistics.

Intervention Development
The findings from the co-design stage were subsequently used to develop a prototype intervention incorporating an app, SMS text messaging, social media, and videos. The intervention development was an iterative process, with meetings held among the team members to decide the final features and functionalities. Not all ideas could be adopted because of various reasons, such as time, funding, and technology constraints. We used a combination of building new functions (app) and using existing functions (Facebook page and SMS text messaging).

The intervention was grounded in behavior change theory. The Health Belief Model was used to underpin the app portion of the intervention. The Health Belief Model is considered to be well suited to mHealth interventions with use of the cue to action component [33]. The basic constructs are perceived threat of illness, perceived benefits of health behavior change, perceived barriers to change, cues to action, and self-efficacy [34]. Behavior change techniques were used to formulate SMS text messages. The SMS text messages were coded for behavior change techniques by 2 coders (Sam McCrabb and SJP) using behavior change technique taxonomy (version 1) [35] and the process outlined by Michie et al [36]. Of the 2 coders, 1 was experienced in coding behavior change techniques (Sam McCrabb) and the other was a PhD student (SJP). Disagreements were resolved through discussion and key messages adapted to include further effective behavior change techniques.

Key messages were developed on health topics identified from the focus groups and interviews. Content was formulated from publicly available evidence-based health resources. Key messages were adapted to SMS text messages, small pieces of written information for the app, and Facebook posts.

The prototype intervention included an app, videos, Facebook page, and SMS text messaging (Textbox 1).
Textbox 1. Components of the prototype intervention.

App
- A web-based prototype app was developed. Rapid iterative cycles between the app developer and research team were used to refine the design. An Aboriginal graphic designer developed graphics for each module and logo.

Videos
- A total of 12 short videos were captured on a Canon camera. All presenters were health professionals from participating sites or contacts of the research team. Short scripts were provided to health professionals based on key messages. Staff were encouraged to use their own knowledge and expertise on each topic. Videos were filmed by a videographer and professionally edited. Captions were completed by Rev, and voiceovers were completed by 2 Aboriginal researchers (BH and NS). The videos ranged from 112 to 300 seconds in length. Vimeo was used as the platform to host the videos.

Facebook page
- A Facebook group was developed and administrated by 2 Aboriginal researchers (BH and NS). Both researchers were regular Facebook users and had significant networks and knowledge of Aboriginal and Torres Strait Islander organizations, events, and health services. Key messages were predeveloped in text and video format. Other content shared was decided by the administrators, including sharing posts from their personal accounts if they were suited to the broad aim of the intervention.

SMS text messaging
- SMS text messages were developed based on the processes described by Abroms et al [37]. Steps include choosing a behavior change goal, choosing communication objectives and behavioral techniques, designing a framework, and writing an SMS text message library [37]. SMS text messages were written to allow tailoring using the mother’s and child’s names, child’s age, and topic interest of the mother. Tailoring SMS text messages around the timing of key behaviors, such as after a baby is born, can improve saliency and likelihood of behavior change [38]. SMS text messages were written by an Aboriginal researcher (BH) and a PhD student (SJP). A web-hosted SMS text messaging server (SMS Express) will be used to send all SMS text messages.

Results

Overview
A total of 42 participants were recruited to the study: 31 mothers and 11 health professionals. Demographics and cultural characteristics of mothers are presented in Table 1, and demographics of health professionals in Table 2.
Table 1. Demographic and cultural characteristics of mothers (N=31).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD; range)</td>
<td>31.17 (7.69; 19-50)</td>
</tr>
<tr>
<td><strong>Indigenous status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Aboriginal</td>
<td>21 (68)</td>
</tr>
<tr>
<td>Torres Strait Islander</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Nonidentified</td>
<td>7 (23)</td>
</tr>
<tr>
<td>Did not answer</td>
<td>1 (3)</td>
</tr>
<tr>
<td><strong>Identified with an Indigenous community, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>25 (81)</td>
</tr>
<tr>
<td>No</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Unknown</td>
<td>4 (13)</td>
</tr>
<tr>
<td>Did not answer</td>
<td>1 (3)</td>
</tr>
<tr>
<td><strong>Maintain cultural connections at home, yes, n (%)</strong></td>
<td>25 (81)</td>
</tr>
<tr>
<td><strong>Ways of connecting to culture, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Music or dance</td>
<td>19 (61)</td>
</tr>
<tr>
<td>Storytelling</td>
<td>19 (61)</td>
</tr>
<tr>
<td>Indigenous television</td>
<td>18 (58)</td>
</tr>
<tr>
<td>Art</td>
<td>15 (48)</td>
</tr>
<tr>
<td>Food</td>
<td>14 (45)</td>
</tr>
<tr>
<td>Indigenous internet sites</td>
<td>10 (32)</td>
</tr>
<tr>
<td>Indigenous newspapers</td>
<td>7 (23)</td>
</tr>
<tr>
<td>Traditional medicine</td>
<td>6 (19)</td>
</tr>
<tr>
<td>Indigenous radio</td>
<td>5 (16)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (3)</td>
</tr>
<tr>
<td><strong>Family members from Stolen Generations(^a), n (%)</strong></td>
<td>6 (19)</td>
</tr>
<tr>
<td>Yes</td>
<td>6 (19)</td>
</tr>
<tr>
<td>No</td>
<td>12 (39)</td>
</tr>
<tr>
<td>Unknown</td>
<td>13 (42)</td>
</tr>
<tr>
<td><strong>Education of mother, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Did not finish high school</td>
<td>6 (19)</td>
</tr>
<tr>
<td>High school</td>
<td>6 (19)</td>
</tr>
<tr>
<td>Certificate</td>
<td>10 (32)</td>
</tr>
<tr>
<td>Diploma</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>4 (13)</td>
</tr>
<tr>
<td>Postgraduate degree</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Did not answer</td>
<td>2 (7)</td>
</tr>
<tr>
<td><strong>Currently pregnant, yes, n (%)</strong></td>
<td>1 (3)</td>
</tr>
<tr>
<td><strong>Partner, yes, n (%)</strong></td>
<td>16 (52)</td>
</tr>
<tr>
<td><strong>Number of people living in household, mean (SD; range)</strong></td>
<td>4 (1.31; 2-7)</td>
</tr>
<tr>
<td><strong>Number of children (aged &lt;18 years) living in household, mean (SD; range)</strong></td>
<td>2.39 (1.41; 1-5)</td>
</tr>
<tr>
<td><strong>Smoking status of mother, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Nonsmoker</td>
<td>21 (68)</td>
</tr>
<tr>
<td>Yes, daily</td>
<td>5 (16)</td>
</tr>
<tr>
<td>Characteristics</td>
<td>Values</td>
</tr>
<tr>
<td>----------------</td>
<td>--------</td>
</tr>
<tr>
<td>Yes, at least once a week</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Yes, less often than once a week</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Did not answer</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Number of cigarettes smoked per day (on the days smoking), mean (SD; range)</td>
<td>8.5 (3.21; 4-12)</td>
</tr>
<tr>
<td><strong>Number of smokers in household, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>14 (45)</td>
</tr>
<tr>
<td>1</td>
<td>10 (32)</td>
</tr>
<tr>
<td>2 to 3</td>
<td>4 (13)</td>
</tr>
<tr>
<td>&gt;3</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Child exposure to indoor tobacco smoke, yes, n (%)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Child exposure to outdoor tobacco smoke, yes, n (%)</td>
<td>15 (48)</td>
</tr>
<tr>
<td>Child exposure to tobacco smoke in the car, yes, n (%)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

**Table 2.** Demographics of health professionals (N=11).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Health service type, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Aboriginal medical service</td>
<td>6 (55)</td>
</tr>
<tr>
<td>NSW** Health service</td>
<td>5 (45)</td>
</tr>
<tr>
<td>Sex: female</td>
<td>11 (100)</td>
</tr>
<tr>
<td><strong>Indigenous status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Aboriginal</td>
<td>4 (36)</td>
</tr>
<tr>
<td>Torres Strait Islander</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Nonidentified</td>
<td>7 (64)</td>
</tr>
<tr>
<td><strong>Role at health service, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Registered nurse</td>
<td>7 (64)</td>
</tr>
<tr>
<td>Aboriginal health worker</td>
<td>3 (27)</td>
</tr>
<tr>
<td>Senior family health practitioner</td>
<td>1 (9)</td>
</tr>
<tr>
<td><strong>Number of years at service, mean (SD; range)</strong></td>
<td>12 (8.7; 3-32)</td>
</tr>
</tbody>
</table>

*NSW: New South Wales.*

**Design Characteristics**

We identified six main design characteristics: (1) credibility, (2) Aboriginal and Torres Strait Islander designs and cultural safety, (3) family centeredness, (4) supportive, (5) simple to use, and (6) confidential.

**Credibility**

Mothers talked about the difficulty of finding information on the web that was evidence based. Most of the mothers said that they used Google to find real-time health information for themselves and for their children: "Literally, I Google everything." Many of the mothers said that it can be difficult to know which websites are most up to date and accurate and that it is difficult to find information: "The biggest thing I find on Google, you get everything. You don’t get the ones that are reputable." Another mother said, “I’m finding you’re having to like scroll, scroll, and scroll to try and find that information.” Mothers said that they want current health information from reputable health professionals and organizations, including “useful websites links.” Health professionals talked about the importance of credible health information to improve health literacy: “I think lack of knowledge that they are so sick. Recognizing the signs of illness that can lead to them being really, really [sick].” This highlighted why it is important that all content included in the prototype intervention be sourced from credible evidence-based health resources and broken down into palatable small chunks with links to further information.
Aboriginal and Torres Strait Islander Designs and Cultural Safety

Most of the mothers said that Aboriginal designs, language, and representation were important for engagement. A mother said, “I think if it had Aboriginal designs that would be really good because if I download an app and it doesn’t have the look, like being culturally aware [I don’t use it].” Another mother said, “Don’t make it black and white, it’s got to be like colorful.” A mother spoke about the intervention needing Aboriginal representation in images and videos: “If it’s going to be an Aboriginal app, I think you have to have Aboriginal people.” Another mother discussed using an app for quitting smoking that was not representative of Aboriginal people: “It was easy to use, but I couldn’t relate to it...didn’t seem like it was aimed at Blackfellas even though we thought it was.”

It was evident from the mothers’ experiences of racism that the intervention needed to be centered in culturally safety. Some mothers talked about feeling fearful and judged when seeking health care. A mother said, “Being an Aboriginal mum especially, I was just worried about DoCS [Department of Child Services]. Like whether they could see if I was handling having two children on top of my own family breakdown. Like my mum’s kids are in DoCS. So that’s what my biggest fear was.” Other mothers expressed feeling judged about certain health behaviors and topics, and a mother said, “The biggest thing is why people do hide it [smoking], because they don’t want to be judged. They don’t want to hear all that stuff.”

To center cultural safety in the intervention, all aspects of the intervention were codeveloped by Aboriginal people: the research was governed by an Aboriginal advisory board and coled by an Aboriginal academic (KH); 4 of the 8 members of the research team are Aboriginal; an Aboriginal graphic designer designed the module icons and logo; Aboriginal researchers were administrators of the Facebook page and shared cultural links, events, activities, affirmations, and images; an Aboriginal videographer filmed all the videos; Aboriginal health professionals presented in the videos; an acknowledgment of Country and a welcome message by an Aboriginal researcher was placed on the main page; and all content was cowritten by Aboriginal researchers.

Family Centeredness

It was decided unanimously that the intervention should include content for both mother and child. A mother said, “Is this just for children’s health? Because I feel like it should incorporate the mother’s health too.” The mothers asked for information on “things to do with our kids,” and “stuff for us women too. Pap smears and stuff like that.” Many of the mothers and health professionals suggested that the intervention needed to encompass the entire family, including the extended family. A health professional said, “Put the main focus on the child and then how their [family] health affects the baby’s health,” and a mother said, “I think a family app would be really good. Like, I know my husband, he’s never been around babies.” Some participants talked about how other family members help bring up children: “It’s nothing to see an aunt bringing up a child, or a grandparent or a sister” [health professional]. Family centeredness in the intervention was therefore conveyed through messaging that families are the most important role models for jarjums (an Aboriginal word meaning children) across modules and functions. Links to websites, events, and health information for partners and other family members were included.

Supportive

Most of the mothers and health professionals indicated that it was important that the intervention promoted positive self-esteem and well-being of mothers. A health professional said that the intervention should give new mothers “understanding [of] how tired you are going to be, and it’s okay, ask for help, everyone feels like that but you’re not failing or not doing something wrong.” A mother suggested that we include “some sanity sayings or something like that, or some little sage advice from mums that have been there, done that before, that’d be really helpful,” and another mother said that the intervention could be “like a reassurance type thing.” Mothers and health professionals recognized that motherhood can be “totally exhausting” [health professional] and challenging at times. A mother described the initial period after coming home from hospital: “I didn't know what to do with him. What do I do with this kid? I was lost.” To create an intervention that was supportive of motherhood and of Aboriginal and Torres Strait Islander women, positive and affirming messages were posted on Facebook, sent through SMS text messages, and included in the app. Links on where to seek help for mental health concerns were included.

Simple to Use

Mothers and health professionals recommended that the intervention be intuitive, use simple language, and have few technical barriers. Some of the mothers talked about trying to use other health apps; however, they were unable to do so because of technological challenges. For example, a mother said, “It was just too hard to log in and get started so I gave up or just called someone.” Many of the mothers and health professionals emphasized that the language used in the intervention needed to be non-jargon. A mother said, “Don’t put it in a textbook. Because I’m telling you, if my family member downloaded that and it was a textbook way, they would be like—No.” Another mother said that the content should be “just little pieces of information...then links to the bigger pieces.” We aimed for simple, intuitive app design and used other mobile functions commonly used by mothers (Facebook and SMS text messaging). To ensure that the intervention was simple and easy to use, health information was presented in short key messages with links to websites for further information. All key messages were written to be at an 8th grade reading level using the Flesch–Kincaid Grade Level Test as recommended by Abrons et al [37].

Confidential

Mothers and health professionals talked about the importance of confidentiality. Health professionals focused on confidentiality in the health care setting and the complexities for some staff regarding knowing patient health details. A health professional said, “There are big things surrounding our health services confidentiality. People don’t know or want to know what other people’s business is.” Some of the mothers spoke
about confidentiality; regarding being anonymous when communicating with other mothers or health professionals in a hypothetical mHealth intervention, a mother said, “Oh God, yeah. I’d ask an anonymous person on a phone. Rather than ask the doctor face to face.” Other mothers were happy to not be anonymous: “It wouldn’t bother me having my name because it would just be, this is my experience, and it is what it is. But I would understand if some women didn’t.” To ensure that women can choose to remain anonymous and keep their information confidential, the intervention design meant that no personal data were collected in any part of the intervention, other than a mobile number for the SMS text messaging component. Joining the Facebook group is an optional part of the intervention.

**Content Modules**

Most of the mothers and health professionals suggested that the intervention needed to cover a wide range of health topics for both the mother and child. Health topics identified in the data included *pains after birth, breastfeeding, normal speech for toddlers, signs of autism, earaches, behavior, rashes, high temperatures*, and *coughs*. Similar topics were grouped by the research team and combined into 6 key content modules for women’s health and 6 key content modules for children’s health. For example, *birth, reproductive health, urinary leakage*, and *pap smears* became *Women’s business*. All health topics captured in the interviews and focus groups were included in the intervention within a module on the app, SMS text messages, or through Facebook posts. Health modules for women included *Smoke-free families, Safe drinking, Feeling good, Women’s business, Eating, and Exercising*. Health modules for children were *Breathing well; Sleeping; Milestones; Feeding and eating; Vaccinations and medicines; and Ears, eyes, and teeth*.

**Features and Functions**

We identified eight features and functions: (1) content feed, (2) social connection, (3) diary and storage of health information, (4) local context, (5) reminders, (6) rewards, (7) talk with health professionals, and (8) use of videos.

**Content Feed**

A content feed was chosen to be a feature of the intervention based on the mothers’ current mobile phone use. During the card-sorting activity, most of the mothers reported scrolling the content feed on Facebook numerous times per day. Of the 13 women who were asked how many hours per day they used Facebook, 12 (92%) reported using it >4 hours per day. When asked what kept them going back to Facebook, a mother responded, “The content keeps changing.” Mothers frequently talked about watching photo and video stories that were uplifting, funny, or motivating on Facebook. They talked about using Instagram and Snapchat, too, although less frequently. The intervention was therefore designed to include a Facebook page with daily posts covering a variety of health content.

**Social Connection**

Mothers talked about the social connection and learning from other women when becoming a mother, including from their “mum,” “mother-in-law,” and “girlfriends.” The importance of positive relationships when first becoming a mother was well recognized by health professionals as well as mothers. It was acknowledged by many of the mothers that some new mothers “don’t have a big support network.” A mother described mothers at playgroup being “more like a family to each other.” Some of the mothers said that connecting to other mothers would be helpful because they may be going through the same situation or challenge: “Yeah [I would like to chat with mums in the intervention] because they might have experienced something that I’m starting to experience.” Some of the mothers talked about the possibility of meeting up with mothers outside of the intervention: “It’s hard to meet people...[could there be] like a mums and bubs [babies] thing [as part of the intervention],” and another mother said, “Say, if I needed to ask them a question or something that I wouldn’t want to write on Facebook [I would like to meet up with them in person].” Another mother identified that connection is important for mental health: “When they [new mothers] don’t have anybody, depression kicks in.” The Facebook page was designed to make it easy for mothers to connect and share stories and ideas. Discussion points were created to be posted on the Facebook page to facilitate discussion; for example, “Tell us how you engage your jarjums in cooking or take a pic or video of your deadly (great or excellent) li’l chef in the kitchen.”

**Diary and Storage of Health Information:**

A feature that enabled users to store specific information about a child’s health received mixed responses. Some of the mothers thought that having their child’s health information on hand would be of practical benefit when attending medical appointments: “Like a diary section...I found, when [my child] was sick I started recording when I gave the medication, those sorts of things. That’d be good to have an app when you go into the hospital, you go, this is his recordings.” Another mother said, “So they [health professionals] could just add in medication, add in reports...it’d be good because like [the health service] is only open during the week. Usually, like on the weekend, I’d have to go up to the hospital...So it would be good if there was information like after the visit. Because you don’t always take everything in. It goes right over your head.” Other mothers and health professionals thought there would be confidentiality concerns. Because of the confidentiality concerns raised in the co-design process, a diary feature was not included, although it may be considered as an optional feature in future iterations.

**Local Context**

Many of the mothers and health professionals spoke about the uniqueness of their community and said that the intervention needed to be relevant to each community, including language and environment (eg, coastal and desert), as well as health services and other resources. A mother suggested, “You could put in your postcode, location, or area or something and then it could be localized,” and a health professional said, “The contact numbers, if they can’t get into emergency, the [local] health line numbers where they can get a bit of advice would be handy on there as well.” The intervention included phone numbers of local health services for each community in the app, and Facebook posts were designed promoting local health services, events, organizations, and languages.
Reminders
Many of the mothers talked about the usefulness of SMS text messaging reminders from their health services for appointments, and they said that reminders for other areas of health care would be useful too. A mother said, “I would probably like all of them [milestone reminders]. I’d like the whole lot, make sure I’m not missing anything.” Another mother said, “If someone notified me on this app that I’m due for a [pap smear] or something like that, I would like being reminded of things like that.” Most of the mothers said that they would prefer reminders through SMS text messages rather than a push notification from an app because they could go back to the message and reread it. For the intervention, SMS text messages were developed covering a range of reminders, including vaccinations, developmental milestones, check-ups, smoking quit date, exercise, and eating well. Reminders about local health initiatives and events were also created for posting on the Facebook page.

Rewards
The mothers talked about rewards and incentives from health programs and services increasing their motivation. They talked about material rewards such as “shirts,” “caps,” and “supermarket vouchers,” as well as social rewards, including “comments” and “likes” on social media and “clapping” and “cheers” on health apps. The mothers who were asked about receiving rewards for a variety of health behaviors were unanimous in their opinion that rewards were enjoyable and motivating. In the intervention, weekly competitions were created for posting on the Facebook page involving mothers sharing a picture of a health activity; for example, active play or exercising with their children. Prize draws were also incorporated into the intervention for those who participated in the competitions.

Talk With Health Professionals
Some of the mothers suggested that being able to communicate with health professionals using SMS text messages or a live chat function would be beneficial. Some of the mothers said that this function would be useful to confirm whether they required face-to-face health care and for reassurance. A mother said, “Sometimes you don’t know if you should go up there [health service] or not, so you could kind of message and say, ‘Hey, this is what’s happening...is it worth coming up or is it just a viral thing going around?’” Another mother said, “I know a lot of women are just like, ‘What do I do?’ So just having that reassurance I suppose online.” Another mother suggested that it would be helpful to be able to ask health questions anonymously: “The option to be anonymous or not known by people [health professionals] would be handy I guess for more embarrassing health concerns.” Mothers living in rural areas mentioned being anonymous more often in the discussions. Although it was suggested, facilitating a chat with health professionals directly was out of the scope of the current prototype because of cost and resources. Telephone numbers for national, state, and local health services were listed in the app to enable users to connect with health professionals, if needed, regarding the questions they might have.

Use of Videos
Most of the mothers reported during the card-sorting activity that they frequently watched short videos on social media and YouTube. A number of mothers and health professionals advised us that videos and images may be more accessible and preferable for some mothers. A health professional said, “Videos, everyone can watch a video and understand.” Therefore, a video for each health module was developed for the intervention. Each video was stored in the app and added to the Facebook page. Additional health videos from external sources were also able to be shared on the Facebook page.

Final Prototype
The final mHealth intervention, named Growin’ Up Healthy Jarjums, aimed to improve health knowledge and health behaviors, along with providing access to health services. The intervention comprises 3 delivery modalities: app, SMS text messaging, and Facebook page.

App
The app is a central place for users to access all content. The app is primarily for the user who wants in-depth information and has the necessary digital device, internet connection, and literacy skills to access it. It is designed to allow the user to navigate to the topic of interest; for example, exercise, where they will find small amounts of written information, videos, links to websites, and useful contacts. The user may choose to access any topic, in any order, and consume as much information as they like.

The app has four menu screens: (1) home screen, (2) women’s health, (3) children’s health, and (4) contacts (Figure 1). The home screen includes four buttons: (1) My Health, (2) Jarjums Health, (3) Facebook Page, and (4) Contacts. The user may click on a button to move to the next screen or scroll down to access the embedded Facebook content feed. The embedded Facebook content feed allows the user to remain in the app and read the posts, but to comment or like a post, the user needs to access the Growin’ Up Healthy Jarjums Facebook page. An acknowledgment of Country and a spoken welcome message are also included on the home screen. The women’s health (My Health) menu page includes six buttons, one for each of the women’s health modules: (1) Smoke-free families, (2) Safe drinking, (3) Feeling good, (4) Women’s business, (5) Eating, and (6) Exercising. The Jarjum’s Health menu page has the same layout, including six buttons for the children’s health modules: (1) Breathing well; (2) Sleeping; (3) Milestones; (4) Feeding and eating; (5) Vaccinations and medicines; and (6) Ears, eyes, and teeth. Each module, for example, Breathing well, includes (1) Key messages incorporating perceived threat of illness and benefits of changing health behavior; (2) Tips to address barriers to change through reassurance and credible advice; (3) cues to action; for example, “Each time jarjum sees a nurse or GP ask them to have a quick look in bub’s ears to check if there is any infection”; and (4) links to further information, including skills and activities; for example, exercises and healthy recipes to support self-efficacy. The information is presented using small chunks of written information and videos using the same layout in each module.
**SMS Text Messaging**

Alongside the app, the prototype included an SMS text messaging library comprising 112 SMS text messages (Table 3). The SMS text messaging component allows users access to health information regardless of mobile phone type, Wi-Fi access, or digital literacy. The SMS text messages covered the content topics identified by the participants. The SMS text messaging portion of the program is 1-way (unidirectional), other than 3 SMS text messages developed for users who indicate that they want to quit smoking when registering for the program. In total, 23 behavior change techniques from 15 behavior change clusters were incorporated in the SMS text messages (Multimedia Appendix 1).
Table 3. Example SMS text messages developed for the Growin’ Up Healthy Jarjums modules.

<table>
<thead>
<tr>
<th>Module</th>
<th>Example SMS text message</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Women’s health</strong></td>
<td></td>
</tr>
<tr>
<td>Smoke-free families</td>
<td>Text4jarjum: Giving up the smokes is the best thing you can do for your health. Be a role model and be smoke free. Get support from Quitline 13 78 48 or a doctor and quit for good!</td>
</tr>
<tr>
<td>Safe drinking</td>
<td>Text4jarjum: While under the influence of alcohol, people can make less safe decisions about their jarjums. Check out ‘Safe drinking’ for tips to set limits.</td>
</tr>
<tr>
<td>Feeling good</td>
<td>Text4jarjum: You’re probably not getting much sleep right now. Try to make time for yourself, ask for support from family &amp; friends, and nap when bub does. If you feel that you are not coping, talk to your doctor or midwife. There is help.</td>
</tr>
<tr>
<td>Women’s business</td>
<td>Text4jarjum: Be kind to yourself. Your body has gone through some big changes during and after birth. It will take time to bounce back. Whether you had a caesarean or vaginal birth, both may require rest &amp; time for recovery. Here’s what to expect after birth.</td>
</tr>
<tr>
<td>Eating</td>
<td>Text4jarjum: The Australian Breastfeeding Association has some useful tips on nutritional needs for breastfeeding mums.</td>
</tr>
<tr>
<td>Exercising</td>
<td>Text4jarjum: Any amount of movement is good for you. Start by doing a little, and gradually build up. You could start with a walk around the block a few times a week and then gradually increase.</td>
</tr>
<tr>
<td><strong>Children’s health</strong></td>
<td></td>
</tr>
<tr>
<td>Breathing well</td>
<td>Text4jarjum: A cough is often caused by a cold. Usually, a cough gets better on its own and is not serious, but if your child has a cough that doesn’t go away after TWO weeks, or if you are concerned sooner – see your doctor or child health nurse.</td>
</tr>
<tr>
<td>Sleeping</td>
<td>Text4jarjum: A routine that includes relaxing time like bath, book, a gentle song before bed and a regular bedtime each night can help your child settle better.</td>
</tr>
<tr>
<td>Milestones</td>
<td>Text4jarjum: Playgroups, day care and pre-school are great places for jarjums to play and develop. Contact your AMS (Aboriginal Medical Service) or health nurse and find out what’s on.</td>
</tr>
<tr>
<td>Ears, eyes, and teeth</td>
<td>Text4jarjum: Ear infections are really common and can cause long term hearing loss if not treated. Often there are no signs. Ask your doctor to have quick look in [insert child name] ears each visit to make sure there is no infection.</td>
</tr>
<tr>
<td>Vaccinations and medicines</td>
<td>Text4jarjum: Immunising [insert child name] is a safe and easy way to keep jarjums healthy and prevent disease. To check that [insert child name] is up to date with immunisations click here.</td>
</tr>
<tr>
<td>Feeding and eating</td>
<td>Text4jarjum: It’s recommended you breastfeed exclusively until [insert child name] starts solid foods at around 6 months of age. Keep breastfeeding until at least 12 months and beyond.</td>
</tr>
</tbody>
</table>

**Facebook Page**

The final modality included in the prototype was the Facebook page. The purpose of the Facebook page was to create community and connection, allow 2-way communication, and use a platform that is highly popular among users. Daily content was designed to be added to the Facebook page, including (1) links to reliable health websites, (2) activities for families, (3) weekly competitions, (4) key messages (written and video), (5) events in the community, and (6) supportive affirmative posts. The page was administrated by 2 Aboriginal team members (NS and BH), who shared posts relevant to their community and region. The Facebook page was embedded into the main screen of the app; it could also be accessed through Facebook. Examples of posts are presented in Figure 2.
Discussion

Principal Findings

We codeveloped a prototype mHealth intervention focused on the knowledge of mothers of young Aboriginal and Torres Strait Islander children. The aim of the intervention was to improve health knowledge, health behaviors, and access to health services. The final prototype incorporates 3 modalities—app, SMS text messaging, and Facebook page—and includes a range of health topics. In addition, it is centered on being supportive of mothers and culturally safe.

The modality choices were based on a few factors: (1) early discussions with mothers and health services about the need for an app that is culturally relevant and safe, (2) evidence suggesting that SMS text messaging is the most effective mHealth function for health behavior change, and (3) findings from focus groups and interviews indicating that Aboriginal and Torres Strait Islander women were high users of Facebook and SMS text messaging. As suggested in a recent pilot study of a smartphone app with Aboriginal Australians, a one app fits all approach is unlikely to be successful [24]. Using mHealth modalities commonly used by the target group to deliver a health intervention may appeal to more families.

Strengths and Limitations

The first limitation of this research is that it was initiated by a research institution rather than by the community itself. True co-design should begin with completing a needs assessment with communities to see what the health priorities and potential solutions are for that community [38]. This is well described in a New Zealand co-design study [29,40]. To ensure that adequate time and resources are available for relationship building and needs assessment, both should be specified in protocols and funding applications so that sufficient budgets and time frames are allocated. Second, although the intervention covers a range of topics in brief, it does not cover any topic in depth. Although an mHealth intervention with wide-ranging topics seems to be preferred by participants, this may dilute the impact of the intervention on any one risk behavior. Providing links within the Growin’ Up Healthy Jarjums intervention to specific mHealth interventions for target behaviors may overcome this limitation by providing tips for more intense behavior change for those people who are ready to change. Third, because the participants were from only 3 NSW communities, the intervention may have limited generalizability in other Aboriginal and Torres Strait Islander communities. Aboriginal and Torres Strait Islander communities are made up of >250 language groups in which there is great diversity. If this intervention is to expand to other communities, systematic adaptation of the intervention would need to be carried out to ensure that the intervention is suitable to the context of each community [41].

A key strength of this study is that Aboriginal researchers (BH, NS, and BL) led engagement with participants and community organizations. Understanding the importance of trusted and strong cultural relationships, we only engaged with communities that the Aboriginal researchers had a relationship with, which likely resulted in trust as well as interest in participating in this study. Another strength of this study is the thorough reporting of the co-design processes. Inadequate reporting of intervention development was identified as a weakness in a recent systematic review on mHealth development [33]. An additional strength is the involvement of primary health services and professionals. A recent review on health promotion programs in Aboriginal...
communities highlighted that an important consideration is to partner with primary health care services because they are well placed with frequent patient contact, health expertise, and often intricate knowledge of the community [42]. A final and important strength is that we developed a flexible portal for ongoing development and enhancement. The COVID-19 experience has reinforced how important it is to have alternatives to face-to-face health care. Useful additions in future iterations of this mHealth intervention might include development of a flexible platform suitable for inclusion of initiatives inspired by the COVID-19 pandemic, such as subsidized telehealth and videoconferencing. There are also opportunities to develop content on this platform in Aboriginal and Torres Strait Islander languages to better suit users.

**Comparison With Prior Work**

Design characteristics identified in this study, including social connection and family centeredness, reflect Aboriginal and Torres Strait Islander perspectives of health. Connection to family, community, and culture, among other factors, are understood to be equal contributors to health [43]. Arabena et al [44] suggest that community and social connection can ultimately be the health promotion intervention for Aboriginal and Torres Strait Islander communities.

The finding that Aboriginal and Torres Strait Islander women were high users of social media, in particular Facebook, was unsurprising. Aboriginal and Torres Strait Islander health organizations have capitalized on the popularity of Facebook among Aboriginal and Torres Strait Islander people and have been early and adept users of social media for health promotion [25]. An Aboriginal-led social marketing campaign for health promotion, Deadly Choices, has 94,035 Facebook followers, 19,300 Instagram followers, and 9000 TikTok Followers [26,45].

As stated earlier, the methodologies used in this study were based on a co-design study for a health app with Māori and Pacific Islander people [29,40]. There were a number of similar co-design findings. In both studies, participants expressed a holistic view of health and connections to people and place as being central components of health. Participants in both studies talked about a family approach to health, rather than an individual approach, as well as accessible healthy activities in the community. Social support was found to be an important strategy in both studies.

Culture was also identified as important in both studies, although cultural representation may have been a more nuanced finding in the New Zealand study. In our Australian-based study, participants expressed the importance of Aboriginal and Torres Strait Islander representation in terms of designs, colors, images, people, organizations, and safety. Participants in the New Zealand study [29,40] expressed the need to include Māori knowledge, Whakapono (faith and spirituality), and Whakataukī (traditional proverbs), which were to be woven throughout the intervention; for example, the app depicts the completion of challenges as colored footsteps, which is analogous to the journey that the participants’ tūpuna (ancestors) embarked on. There may be differences in participants’ connection to culture. In Australia, up to 1 in 3 Aboriginal and Torres Strait Islander children were removed from their families during the period from the mid-1800s to the 1970s. These children are known as the Stolen Generations [39]. Of the 31 mothers in this study, 6 (19%) reported that they had family members from the Stolen Generations, whereas 13 (42%) were unsure. The effect of the Stolen Generations on the loss of culture is profound [39] and is likely reflected in the findings of this study. This intervention may, in a small way, help to promote culture through links to Aboriginal and Torres Strait Islander organizations, connection to mothers of Aboriginal and Torres Strait Islander children, and culturally safe health information.

**Conclusions**

An mHealth intervention that included app, SMS text messaging, and Facebook page modalities was developed based on co-design findings. The intervention incorporates health behavior change theory, evidence-based information, and the preferences of Aboriginal and Torres Strait Islander women and health professionals. The next step of this research is to assess the acceptability and feasibility of the intervention in a pilot study. The pilot study will be conducted with the Aboriginal Health Services and NSW Health sites that participated in this co-design study. Participating mothers will also be invited to participate in the pilot study. If the Growin’ Up Healthy Jarjums intervention is shown to have adequate acceptability and feasibility, the next phase will be to measure its effectiveness in improving health knowledge and changing health behaviors. Assessing the effectiveness of this intervention will provide valuable evidence for the use of mHealth in improving the health and well-being of Aboriginal and Torres Strait Islander populations and contribute to the evidence for using co-design methodologies, both of which have been highlighted as gaps in the literature [46].

**Acknowledgments**

The authors would like to acknowledge and pay their respects to the Aboriginal and Torres Strait Islander women who shared their views and knowledge in the study. The authors would also like to acknowledge the Aboriginal Community Controlled Health Services that participated in the study (Armajun Aboriginal Medical Service [AMS], Galambila AMS, Awabakal AMS, Kulai Aboriginal preschool, and Muloobinba Aboriginal Corporation) as well as the New South Wales Health sites that participated (Aboriginal Maternal and Infant Health Services located in Inverell, Coffs Harbour, and Newcastle). The authors wish to thank researchers from the National Institute for Health and Innovation, Auckland University, Auckland, New Zealand, for sharing expertise on mobile health and Sam McCrabb for coding behavior change techniques.
Conflicts of Interest

None declared.

Multimedia Appendix 1

Behavior change techniques in SMS text messages.

References


**Abbreviations**

AMS: Aboriginal Medical Service

mHealth: mobile health

NSW: New South Wales
Abstract

Background: Despite experiencing the second-highest rate of HIV incidence in the United States, pre-exposure prophylaxis (PrEP) use remains low among Black women due, in part, to a lack of patients’ awareness and providers’ knowledge.

Objective: Our goal was to design animated educational tools informed by patients and women’s health providers to address these barriers, specifically for women at risk for HIV.

Methods: Two animation storyboards about PrEP for women were created by academic stakeholders (e.g., HIV clinical experts, educators, and HIV peer counselors), one for patients and one for providers. Four focus groups with community members from Baltimore, Maryland and four with women’s health providers (e.g., obstetrician/gynecologists, midwives, nurse practitioners, and peer counselors) at an academic center were conducted to discuss the storyboards. Transcripts were analyzed using conventional content analysis, and themes were incorporated into the final versions of the animations.

Results: Academic stakeholders and 30 focus group participants (n=16 female community members and n=14 women’s health providers) described important themes regarding PrEP. The themes most commonly discussed about the patient animation were understandability of side effects, HIV risk factors, messaging, PrEP access, and use confidence. Provider animation themes were indications for PrEP, side effects, and prescribing confidence.

Conclusions: We created two PrEP animations focused on women. Stakeholder feedback highlighted the importance of ensuring the understandability and applicability of PrEP educational materials while including necessary information to facilitate use or prescribing confidence. Both community members and women’s health providers reported greater use confidence after viewing the animations.

(KEYWORDS: PrEP; animations; education; HIV; prevention; women)
or unknown status sexual partner, a recent sexually transmitted infection, a high number of sexual partners, report inconsistent or no condom use, participate in commercial sex work, live in a high HIV prevalent area, or inject drugs [4].

Despite the availability of PrEP, uptake has been poor among women at high risk [8,9]. Patient-level barriers to PrEP include low self-perceived risk of HIV, limited knowledge, and high perceived cost [10,11]. Provider-level barriers include poor support and infrastructure to provide PrEP, inadequate education, and underestimating patients’ risk [10,11]. Additionally, few PrEP campaigns specifically target women or women’s health providers [10,11].

Computer-based interventions, such as animations, have been associated with decreased high-risk behaviors leading to HIV acquisition [12]. Furthermore, creating multimedia tools for health education with stakeholder involvement has been encouraged to identify specific community needs and ensure effective dissemination [13,14]. Additionally, animations can decrease cognitive overload and increase attention retention and long-term recall [14,15]. Therefore, to address commonly cited barriers to PrEP uptake among women, we sought to create two women-centered PrEP animations, one for providers and one for patients, grounded in Mayer’s [15] Cognitive Theory of Multimedia Learning that posits combined auditory text and visual pictures deepens understanding more than either alone. Below, we describe the animation development process with the participation of community members and women’s health providers from Baltimore, Maryland.

**Methods**

Animations were created and iteratively refined in four phases between January 2020 and December 2020 (Figure 1).

**Figure 1.** Study phases. PrEP: pre-exposure prophylaxis.

<table>
<thead>
<tr>
<th><strong>Phase 1:</strong> Animation Content Creation</th>
<th><strong>Phase 2:</strong> Storyboard Creation</th>
<th><strong>Phase 3:</strong> Focus Group Discussions</th>
<th><strong>Phase 4:</strong> Animation Creation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical HIV prevention experts</td>
<td>Women’s health providers, health literacy experts, HIV peer counselor, patients, branding expert, nurses</td>
<td>Women’s health providers, patients</td>
<td>Animation creation and refinement</td>
</tr>
<tr>
<td>Research team</td>
<td>Research team Animation company</td>
<td>Recorded facilitated group discussions about storyboard script and artwork, selection of professional voice over</td>
<td>Animation creation and refinement</td>
</tr>
<tr>
<td>Key topics to include based on literature, PrEP guidelines</td>
<td>Draft storyboard script of 280 words for a 2.5-minute animation; create artwork (eg, characters, colors, scenery)</td>
<td>Storyboard creation</td>
<td>Animation creation and refinement</td>
</tr>
</tbody>
</table>

**Ethics Approval**

This study was approved by the Johns Hopkins University School of Medicine Institutional Review Board (approval number: IRB00252170).

**Phase 1: Animation Content Creation**


**Phase 2: Storyboard Creation**

Two storyboards (ie, scripts and 2D slides) were created by an animation company (Science Animated, Cotswolds, United Kingdom) based on the outlines written in phase 1. The patient storyboard used simple language to aid low–health literacy populations, had relatable characters, and used positive framing. The provider storyboard assumed prior knowledge of women’s sexual health and was framed for practicality using medical terminology. In addition, we held formative discussions with our academic stakeholders (ie, lay individuals, a patient education professional, a branding director, nurses, and clinical HIV prevention experts), and storyboards were revised based on these initial discussions.

**Phase 3: Focus Groups**

Feedback about the storyboards from a larger audience was gathered. English-speaking women from the community and women’s health providers (eg, obstetrician/gynecologists, midwives, nurse practitioners, and peer counselors) were recruited for focus groups. Fliers were placed in all general gynecology clinics, all academic specialist obstetrician/gynecologists in an academic hospital were emailed about participating, and prior research participants who had agreed to be contacted for future studies were recruited. Purposive sampling was conducted.

After obtaining written consent and collecting demographic information, the focus groups were scheduled. A trained facilitator and logistical coordinator conducted focus groups virtually using a secure videoconferencing platform (Zoom 5.5.4; Zoom Video Communications, Inc). Participants were offered the option to turn off their video and remove their names.
to allow partial anonymity, as all still heard voices. There were 8 focus groups, 4 for community members and 4 for women’s health providers. Each focus group contained 3 to 6 participants and lasted 60 to 90 minutes. A semistructured focus group guide was used to frame the discussions. Near the end of each focus group, participants were asked to rank 6 female-sounding professional voice-over actors who read the same script but may have had differences in inflection, cadence, pitch, or articulation. Participants received a US $25 compensatory gift card.

**Phase 4: Focus Group Analysis and Animation Creation**

The focus group audio was recorded and transcribed verbatim. The transcripts were coded using ATLAS.ti 9.0.3 (ATLAS.ti Scientific Software Development GmbH), and the findings were organized and analyzed using conventional content analysis [19]. Two research team members read the transcripts twice, line-by-line, and prepared memos summarizing their preliminary findings. Next, preliminary codes were derived inductively by highlighting recurring words or statements through an iterative process. Research team members convened multiple times to discuss and compare memos, and revisit emerging themes iteratively. Discrepancies were solved by the principal investigator. Final codes were then assessed for broader concepts to generate themes through subsequent rounds of team discussion. The generation of themes was guided by an adaptation of the Model of Communication and Health Behavior Change by Kincaid [20]. The themes identified in the storyboards (phase 2) and focus group discussions were incorporated into animation prototypes. The highest-ranked voice-over options were chosen. Finally, two 120-second 2D animations were created and iteratively refined by the research team.

**Results**

**Storyboard Creation (Phases 1 and 2)**

Themes that were presented in the initial creation of the storyboard included accurate PrEP information, ensuring an appropriate health literacy level to reflect the target population (eg, proficient level for the provider animation and basic to below basic level for patient animation) and representative graphics/artwork (eg, multicultural characters and scenery). The scripts for the storyboards and animation graphics were refined numerous times by stakeholders.

**Focus Groups (Phase 3)**

A total of 30 participants enrolled in the focus groups (n=16 female community members and n=14 providers, Table 1). Some themes pertained to both animations (messaging, design, background, side effects and risk factors, and perceptions of PrEP access and barriers), while others were only relevant to the patient animation (relatability and applicability of characters and storyline, and PrEP use confidence) or the provider animation (understandability of PrEP indications and prescribing confidence).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Community member (n=16)</th>
<th>Women’s health provider (n=14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>26 (6)</td>
<td>42 (13)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0 (0)</td>
<td>2 (14)</td>
</tr>
<tr>
<td>Female</td>
<td>16 (100)</td>
<td>12 (86)</td>
</tr>
<tr>
<td><strong>Self-reported race, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>12 (75)</td>
<td>3 (21)</td>
</tr>
<tr>
<td>Asian</td>
<td>0 (0)</td>
<td>3 (21)</td>
</tr>
<tr>
<td>White</td>
<td>1 (6)</td>
<td>7 (50)</td>
</tr>
<tr>
<td>Other</td>
<td>3 (19)</td>
<td>1 (7)</td>
</tr>
<tr>
<td><strong>Highest education level, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>7 (44)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>College</td>
<td>6 (38)</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Graduate school</td>
<td>2 (13)</td>
<td>13 (93)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (6)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>8 (50)</td>
<td>6 (43)</td>
</tr>
<tr>
<td>Married/union</td>
<td>7 (44)</td>
<td>7 (50)</td>
</tr>
<tr>
<td>Divorced</td>
<td>1 (6)</td>
<td>1 (7)</td>
</tr>
</tbody>
</table>

Table 1. Sociodemographic characteristics of focus group participants (N=30).
Provider Feedback
Providers’ most common themes included prescribing confidence, indications for PrEP, and side effects. Additional themes and respective quotes are highlighted in Multimedia Appendix 1.

Indications
Providers were surprised by a few of the indications for prescribing PrEP, specifically in relation to the area in which their patients were living. Several of them were surprised to learn that the HIV prevalence in a location was included in evaluating HIV risk and PrEP indications.

So, automatically living in Baltimore, it puts you at a higher risk for HIV, so really understanding that kinda stood out to me.

Side Effects
The level of details on side effects and their timing was a concern for some providers. Providers asked for more clarification on which side effects to expect and a given time frame for each to be included in the animation.

I wanna like know [...] how long is it [resolution of initial side-effects] gonna take? Are we talking about the next day? Are we talking about a month?

Prescribing Confidence
Providing a more detailed step-by-step guide applicable to the flow of a typical clinic encounter, including follow-up steps after prescribing PrEP, was recommended.

I’d first ask myself like, “Hm, this person seems like they’re high risk for HIV.” And then, I would ask myself, “Do I – do I have testing that allows me to firmly confirm or deny the fact that they actually have HIV right now?” [...] And then, I would say to myself, like, “Okay, so I think if they like don’t have a test on file or it’s not that recent, are there any signs that I think that they’re nonetheless actively infected with HIV and I need to test them for that before, you know, starting a conversation about preventing HIV.”

Providers were, overall, confident in their ability to prescribe PrEP based on these steps but expressed doubt about their patient’s desire and ability to comply with extensive follow-up and regular lab draws.

You’re gonna have people who are not gonna wanna come in for a HIV test every three months...I mean, that’s just gonna be a deterrent for people.

Community Member Feedback
Community members’ most common themes were understandability of side effects and risk factors, messaging, PrEP use, PrEP access, and use confidence.

Risk Factors
Community members were surprised by the prevalence of HIV in their community and that they themselves would qualify for PrEP based on the listed risk factors.

I would say it related to me because before then, I never knew about PrEP [...] And I think that it will be not only a big eye-opener for me but for everyone else.

I didn’t know that – um, that Baltimore City, HIV was as high as it is. They had shared it on the news, I think like a week or two ago, and liked it just kinda like caught me off-guard...

Side Effects
Community members were concerned about serious side effects and wanted more information about drug-drug interactions, specifically interactions between PrEP and contraceptive methods.

The stomach and the headaches, [...] that’s kinda common. But like generally, kidney and bone density, that’s not like average things.

Um, you said that it doesn’t affect pregnancy or anything like that, but is there any risk – Like, if I’m on birth control, and I supplement with PrEP, is there any effect there, or they don’t affect each other whatsoever?

PrEP Use
Community members highlighted confusing concepts, including PrEP’s ability to prevent HIV and the logistics of PrEP follow-up that would need to be clarified to encourage use.

How long does it last? Is it like a shot? Well, I know it’s like a pill, but like how long does it last? Like do you have to take one every day, every week, once a month?

Messaging
Overall, community members reported that the language and the messaging were appropriate for all education levels.

So, if we can clear it up that – Yeah, condoms alone do prevent HIV acquisition, but it’s much more effective if you use PrEP. And if you do both together that’s even better. Um, so maybe there is a way that we could kind of, like, get that message across.

PrEP Access
Community members thought use confidence would be impacted by information about PrEP access, especially insurance coverage. They wanted the animation to convey that it was easy to take PrEP via the step-by-step guide, which motivated participants to recommend PrEP to their peers.

I think the most surprising thing for me is that there is an option for people without insurance [...] I don’t know how true that is. Because y’all always say that, but they’ll be, like, “Yeah. There’s an option. You can take off 10 percent.” That’s not enough.

Use Confidence
Overall, there was a potential for greater use confidence after viewing the storyboard. Community members expressed that...
the storyboard motivated them to read more about PrEP and initiate a discussion with their provider.

But now, after this focus group, I’m more interested because it was kinda well-explained. I will do my own research on like the bone density and the kidneys and the side-effects, but I think after this focus group that, uh, it’s something that I will have a conversation with my doctor about.

...But, yes, I would. It's very simple, it's appealing. Um, if it's 90% accurate plus on top of a condom, um, especially if you have multiple partners. Why not?

**Animation Creation (Phase 4)**

Stakeholder feedback from all phases led to clarifying language modifications and additional detail to describe the background, risk factors, indications, and steps for prescribing or accessing PrEP. Specifically, the provider animation was modified to clearly delineate prescribing steps. Details about side effects and interactions were added to the patient animation. The importance of evaluating HIV risk and PrEP eligibility according to risk factors, including geographic HIV prevalence, was explained better. Finally, design changes to the characters were made to make them more relatable (Multimedia Appendixes 2 and 3).

**Discussion**

**Principal Findings**

Two educational animations to facilitate learning about HIV prevention and PrEP for female patients and providers were created using a user-centered approach. There were some similar themes both community members and women’s health providers wanted to highlight in the animations that included a clear demonstration about indications for PrEP, addressing barriers to PrEP use, and providing step-by-step guides to accessing or prescribing PrEP. These themes were considered the most important for both patients and providers to increase PrEP awareness and uptake among at-risk women. In addition, there was greater use confidence after viewing the storyboards.

**Comparison to Prior Work**

Although we did not test the “real-world” effectiveness of the final PrEP animations in this formative study, in general, animations have been found to be effective in increasing health information recall [12,16,21]. One study used a 2 x 2 factorial design among patients with different health literacy levels to determine which features of animations improved health information recall and attitudes [21]. They found that spoken animation significantly improved recall of health information compared to written messages among low-literacy participants ($P=.02$). Additionally, there was no differences in health information recall between high- and low-literacy participants after exposure to spoken animation ($P=.12$). Furthermore, a meta-analysis demonstrated that technology-based HIV prevention interventions have been proven to be at least as efficacious as human-delivered interventions in reducing high-risk sexual behaviors [12].

**Strengths and Limitations**

A notable strength of our study is the user-centered approach with key stakeholders, which has been proven to foster stronger relationships between researchers and the community [22]. This may allow greater dissemination among at-risk women and women’s health providers. However, there are limitations. First, the generalizability of these findings to other locations may be limited. We recruited providers from a single tertiary care center, who do not represent all health care providers in different settings. However, our community members reflect women most impacted by the HIV epidemic and that share similar characteristics. Second, although we collected data until we thought that theme saturation was reached, the sample of participants was small and additional themes might have been missed. Third, the animations did not provide exhaustive information and were not tested for effectiveness. However, the purpose of the animations will be to facilitate a discussion between female patients and women’s health providers as an adjunct to routine sexual health care. Additionally, we do not expect our animations to be less effective than other human-delivered interventions, as existing data has shown that technological interventions are effective [12].

**Conclusion**

To increase the use of PrEP in women who live in communities with high HIV risk, the dissemination of information regarding its use in a relatable, accessible, and applicable way is vital for both patients and providers. Therefore, we included stakeholders in creating short educational PrEP animations. Stakeholders highlighted important issues to them, which included identifying individuals that qualified for PrEP, delineating key steps in accessing or prescribing PrEP, and addressing barriers to PrEP. As a result, there was greater use confidence for community members and women’s health providers after viewing the storyboards. Future research is planned to evaluate the effectiveness of the animations to increase PrEP awareness and uptake among women who are at substantial risk for HIV.

**Acknowledgments**

The authors thank the focus group participants and Rosemary Ramroop, Janelle Ramroop, Gregoria Valdez, Tom Bauer, and Tanique Bennett for reviewing early versions of the focus group guides and animation storyboards.

**Authors’ Contributions**

AMY was responsible for facilitating the focus groups, data analysis, data interpretation, modification of scripts for the animations, and drafting of all manuscript sections. TF was responsible for facilitating the focus groups, data analysis, data interpretation,
creating and modifying scripts for the animations, and drafting all manuscript sections. OO was responsible for facilitating and recording the focus groups, study coordination, participant follow-up, and editing and reviewing the manuscript. EG was responsible for analyzing and interpreting the data, modifying the animation scripts, and editing and reviewing the manuscript. JC posed the research question for this manuscript and was responsible for data interpretation, analysis, editing, and manuscript review.

Conflicts of Interest
JC’s institution receives research funding from Gilead Sciences.

Multimedia Appendix 1
Supplemental Tables 1 and 2.
[DOCX File, 19 KB - formative_v6i7e33978_app1.docx]

Multimedia Appendix 2
Provider animation.
[MP4 File (MP4 Video), 6960 KB - formative_v6i7e33978_app2.mp4]

Multimedia Appendix 3
Patient animation.
[MP4 File (MP4 Video), 7944 KB - formative_v6i7e33978_app3.mp4]

References


5. Guideline on when to start antiretroviral therapy and on pre-exposure prophylaxis for HIV. World Health Organization. 2015. URL: http://apps.who.int/iris/bitstream/handle/10665/186275/9789241509565_eng.pdf [accessed 2021-06-14]


Abbreviations

PrEP: pre-exposure prophylaxis
Adaptation of a Problem-solving Program (Friendship Bench) to Treat Common Mental Disorders Among People Living With HIV and AIDS and on Methadone Maintenance Treatment in Vietnam: Formative Study

Ha V Tran1, MSc, MD; Ha T T Nong1, PhD; Thuy T T Tran2, PhD; Teresa R Filipowicz3, MPH; Kelsey R Landrum3, MPH; Brian W Pence3, PhD; Giang M Le4, PhD; Minh X Nguyen5, PhD; Dixon Chibanda6,7, PhD; Ruth Verhey6, PhD; Vivian F Go8, PhD; Hien T Ho9, PhD; Bradley N Gaynes10, MPH, MD

1The University of North Carolina, Vietnam Office, Hanoi, Vietnam
2Faculty of Environmental and Occupational Health, Hanoi University of Public Health, Hanoi, Vietnam
3Department of Epidemiology, University of North Carolina, Chapel Hill, NC, United States
4Center for Research and Training in HIV/AIDS, Hanoi Medical University, Hanoi, Vietnam
5Faculty of Epidemiology Department, Hanoi Medical University, Hanoi, Vietnam
6Department of Community Medicine & Research Support Centre, University of Zimbabwe, Harare, Zimbabwe
7London School of Hygiene & Tropical Medicine, London, United Kingdom
8Department of Health Behavior, Gillings, School of Global Public Health, University of North Carolina, Chapel Hill, NC, United States
9Faculty of Clinical Medicine, Hanoi University of Public Health, Hanoi, Vietnam
10Department of Psychiatry, University of North Carolina, Chapel Hill, NC, United States

Corresponding Author:
Ha V Tran, MSc, MD
The University of North Carolina
Vietnam Office
Room 407 – 408, Block A2
Van Phuc Diplomatic Compound 298 Kim Ma, Ba Dinh
Hanoi, 10000
Vietnam
Phone: 84 24 3211 5839 ext 108
Email: vietha@live.unc.edu

Abstract

Background: The prevalence of common mental disorders (CMDs) among people living with HIV and people who inject drugs is high worldwide and in Vietnam. However, few evidence-informed CMD programs for people living with HIV who inject drugs have been adapted for use in Vietnam. We adapted the Friendship Bench (FB), a problem-solving therapy (PST)–based program that was successfully implemented among patients with CMDs in primary health settings in Zimbabwe and Malawi for use among people living with HIV on methadone maintenance treatment (MMT) with CMDs in Hanoi, Vietnam.

Objective: This study aimed to describe the adaptation process with a detailed presentation of 4 phases from the third (adaptation) to the sixth (integration) of the Assessment-Decision-Adaptation-Production-Topical Experts-Integration-Training-Testing (ADAPT-ITT) framework.

Methods: The adaptation phase followed a qualitative study design to explore symptoms of CMDs, facilitators, and barriers to conducting FB for people living with HIV on MMT in Vietnam, and patient, provider, and caretaker concerns about FB. In the production phase, we revised the original program manual and developed illustrated PST cases. In the topical expert and integration phases, 2 investigators (BNG and BWP) and 3 subject matter experts (RV, DC, and GML) reviewed the manual, with reviewer comments incorporated in the final, revised manual to be used in the training. The draft program will be used in the training and testing phases.

Results: The study was methodologically aligned with the ADAPT-ITT goals as we chose a proven, effective program for adaptation. Insights from the adaptation phase addressed the who, where, when, and how of FB program implementation in the
MMT clinics. The ADAPT-ITT framework guided the appropriate adaptation of the program manual while maintaining the core components of the PST of the original program throughout counseling techniques in all program sessions. The deliverable of this study was an adapted FB manual to be used for training and piloting to make a final program manual.

**Conclusions:** This study successfully illustrated the process of operationalizing the ADAPT-ITT framework to adapt a mental health program in Vietnam. This study selected and culturally adapted an evidence-informed PST program to improve CMDs among people living with HIV on MMT in Vietnam. This adapted program has the potential to effectively address CMDs among people living with HIV on MMT in Vietnam.

**Trial Registration:** ClinicalTrials.gov NCT04790201; https://clinicaltrials.gov/ct2/show/NCT04790201

(JMIR Form Res 2022;6(7):e37211) doi:10.2196/37211

**KEYWORDS**

Friendship Bench; Vietnam; Assessment-Decision-Adaptation-Production-Topical Experts-Integration-Training-Testing; ADAPT-ITT; common mental disorders; people living with HIV; PWH; people who inject drugs; PWID; methadone maintenance treatment; MMT; depression; anxiety; stress disorder

**Introduction**

Injection drug use is the main cause of the HIV epidemic in Vietnam, resulting in a high HIV prevalence among people who inject drugs, 15% in 2017 [1] and 12.7% in 2020 [2]. Antiretroviral therapy (ART) and methadone maintenance treatment (MMT) have been implemented in Vietnam since 2004 and 2008, respectively, and people living with HIV who inject drugs are commonly treated with simultaneous ART and MMT [3,4]. HIV and drug addiction are chronic stressors on mental health, and when combined with many other barriers related to HIV and drug addiction such as stigma and discrimination, the prevalence of common mental disorders (CMDs) among people living with HIV who inject drugs is higher than in the general population [5,6]. CMD is a collective term that refers to a range of depressive, anxiety, and stress-related disorders [7]. Depression is characterized by sadness, low self-esteem, tiredness, lack of concentration, lack of interest, and sleeping difficulties. Anxiety disorders refer to the feeling of anxiety, fear, panic including posttraumatic disorder, and social anxiety [7]. Stress is a state of physical and emotional tension caused by reactions to stressful stimuli from life and the environment [8]. The prevalence of depression and anxiety disorder among MMT clients in China in 2017 was 12.8% and 19.5% [9], compared with only 4.4% and 3.6% in the general population in 2015, respectively [7]. In a study by Levintow et al [10], people living with HIV who inject drugs in North Vietnam had a mild depression rate of 25% and moderate depression rate of 44%. Mughal et al [11] reported that people living with HIV on MMT in Hanoi had CMDs of depression and anxiety and postrumoral stress rate of 17% [11]. Prior research has demonstrated the effectiveness of behavioral and mental health programs among people living with HIV [12], including an evidence-informed program (EIP) to decrease levels of distress among people living with HIV in Thai Lan [13], an adapted life step intervention to effectively improve depression among people living with HIV in Zimbabwe [14], and an adapted CMD program for people living with HIV proved effective in 2020 in Malawi [15]. In Vietnam, we found evidence that family group counseling improved depression symptoms among people who inject drugs in 2017 [16]. Other studies have also recommended the need for CMD programs for people who inject drugs in Vietnam [17]. Problematically, there have been no mental health programs for this population and setting.

**Friendship Bench (FB),** developed by Chibanda Dixon in 2006 [18], is a brief evidence-informed psychological program using problem-solving therapy (PST) and implemented by lay counselors. More than 50,000 people have received counseling in the FB program, making it the largest mental health program integrated into primary health care in Africa, supported by the government and the Ministry of Health of Zimbabwe [18]. FB has been validated through a clinical trial to reduce CMDs among community members in Zimbabwe. In this study, people receiving FB sessions had lower depression and anxiety scores on the Shona Symptoms Questionnaire (SSQ-14) than those who did not receive FB sessions (3.81 and 8.90, respectively). Moreover, the percentage of people in the intervention arm who had depression symptoms on the Patient Health Questionnaire-9 was lower than that in the control arm (13.7% in the intervention arm compared with 49.9% in the control arm) [19]. FB was used in the provision of ART adherence and depression counseling sessions to people living with HIV in Malawi. A total of 501 people living with HIV who had just started ART and had mild depression (Patient Health Questionnaire-9 scores from 5 to 9) were randomized into 2 groups: the intervention group and the control group. The study indicated that people living with HIV in the intervention group had a higher rate of adherence to ART than those in the control group [15]. Recently, FB had been adapted to reduce depression and anxiety among people living with HIV adolescents in a clinical trial in Zimbabwe. The study of this adaptation is ongoing [20]. FB is also currently being adapted for use among people living with HIV who are pregnant in Malawi [21].

FB consists of up to 6 structured 45-minute sessions, and it has been effective in treating CMDs among community members in Zimbabwe [19,22]. The sessions were conducted outside of the primary clinics. Each session followed a PST approach and included steps to determine a manageable problem, choose a problem to solve, make a plan to solve the problem, create a follow-up of intervention implementation, and encourage work on new problems. After completing an individual session, participants are invited to take part in group activities, where people facing similar life challenges share their stories, spend time together,
and work on income-generating activities [22]. The original FB manual has 23 chapters, covering the topics of counseling techniques and session structure, core competencies of counselors, mental illness, HIV and mental health, substance use disorders (SUD), and PST. However, to implement this program for a new population, we first simplified and restructured the program to meet local norms [23], the new target population, and the new study settings [24] via the Assessment-Decision-Adaptation-Production-Topical Experts-Integration-Training-Testing (ADAPT-ITT) model [25].

The adaptation method ADAPT-ITT of the Centers for Disease Control and Prevention (CDC) was developed by Wingood and DiClemente in 2008 as a framework to guide the adaptation of a proven EIP to the specific goals and study participants of HIV prevention [25]. The ADAPT-ITT model has been used to adapt HIV prevention EIP for HIV key populations such as African American couples in the United States [26], people living with HIV in Zimbabwe [27], African American women in the United States [28], and men who have sex with men in Thailand [29] and to address mental and sexual health for young people living with HIV in sub-Saharan Africa [30].

This study is nested within the parent study “Adaptation of the Friendship Bench counseling programme to improve mental health and HIV care engagement outcomes among people living with HIV who inject drugs in Vietnam,” which aimed to test the feasibility, acceptability, and fidelity of the adapted FB in Vietnam [31]. This paper describes the process of FB adaptation to address CMDs among people living with HIV on MMT in some MMT clinics in Hanoi, Vietnam, in 2021, in preparation for a randomized controlled trial (RCT) following the ADAPT-ITT framework and specifically focuses on ADAPT-ITT phases 3 to 6.

Methods

Overview

The ADAPT-ITT model [25] comprises 8 sequential phases. Phase 1, the assessment, was conducted in a previous study [17]. Phase 2, the decision, was conducted by the parent study [31]. This paper describes in detail from the third to the sixth phases of the model: phase 3 (adaptation), phase 4 (production), phase 5 (topical experts), and phase 6 (integration). Phase 7 (training) and phase 8 (testing) will be presented in future papers.

Phase 1: Assessment

This phase involved needs assessment of people living with HIV on MMT for reduction in CMDs. The study conducted 28 in-depth interviews (IDIs) with people living with HIV who inject drugs (n=16, 57%), HIV and MMT providers (n=8, 29%), and health care providers (n=4, 14%) in private clinic rooms in Hanoi, Vietnam in May 2018.

Results of the study showed that both health care providers and people living with HIV who inject drugs believed that people living with HIV who inject drugs were particularly susceptible to CMDs, especially depression, and had a high need for mental health treatment. The study also recommended integrating mental health care into MMT clinics to increase mental health access to people living with HIV on MMT [17].

Phase 2: Decision

This phase involved reviewing evidence-informed CMD programs and deciding which EIP to select for the new target population. Few programs addressing CMDs for people living with HIV exist in low- and middle-income countries, which have limited mental health resources. One of the few evidence-informed CMD counseling programs is the “FB” model, which has extensive validation in low- and middle-income countries [19]. FB helped to reduce CMDs, which was the aim of this study. In addition, FB has also been adapted for people living with HIV in Zimbabwe and Malawi [15,20,21] and may be suitable for people living with HIV on MMT in Vietnam.

Phase 3: Adaptation

In phase 3, stakeholders including people living with HIV on MMT, their family members, and health care providers reviewed the original FB content and implementation methods to answer questions related to who, when, where, and how to deliver FB in Vietnam for appropriate implementation in MMT clinics. The study conducted IDIs with 12 people living with HIV on MMT, 5 family members and caretakers of the patients, and 4 clinic directors. In all, 2 group focus discussions (FGDs) were conducted with MMT health care providers (each group had 5 people).

Inclusion criteria for patients were being aged ≥18 years, willing to participate in the study, and having at least one of the symptoms: depression, anxiety, stress disorders as indicated by the Depression, Anxiety, and Stress Scale-21 items (DASS-21), consisting of 21 questions and 3 items measuring levels of the emotional states of depression, anxiety, and stress. The DASS-21 is based on a dimensional conception of depression, anxiety, and stress as part of the full human experience rather than the categorical approach of either having a disorder or not (as used in the Diagnostic and Statistical Manual of Mental Disorders or International Classification of Diseases) [32]. Although FB was initially tested using the SSQ-14 to screen for CMDs, this tool has not been validated in Vietnam. The DASS-21, however, has been validated in this setting. Originally developed by Lovibond [32] in Australia in 1983 and tested to screen for CMDs in the community, the DASS-21 has been translated and successfully validated for a variety of study participants in Vietnam, including adolescents [33], rural women [34], health care workers [35], and people living with HIV on MMT [17]. The DASS-21 is aligned with the SSQ-14 in terms of screening CMDs, including depression and anxiety disorders in the general population, but is distinguished by having a stress disorder scale. Thus, the study team chose the DASS-21 as a research tool, as it has the advantages of validity in this population, brevity, ease of implementation, and measurement of depression, anxiety disorder, and stress disorder relevant to the purpose of the study. The study will also use suicide risk assessment that was effectively used in the HPTN074 study for people living with HIV who inject drugs in Vietnam [36].
Inclusion CMDs scores were at least a moderate level in which depression subscale score was ≥14, anxiety subscale score was ≥10, and stress subscale score was ≥19. People who were aged <18 years, did not meet the DASS-21 moderate score, had cognitive impairments, or had suicidal thoughts (exclusion, therefore, included referral for standard clinical follow-up care) were excluded from the study. Patients who met the inclusion criteria were invited to participate in an IDI. Family members and caretakers were the patients’ parents, spouses, or main caretakers who already knew the patients’ HIV and MMT status and who were introduced by the patients. Only family members and caretakers who had patients’ consent to approach and who already knew the patients’ HIV and MMT status were invited for IDIs. The inclusion criteria for health care providers were having at least two years of counseling and clinical experience with people living with HIV and people who inject drugs.

IDI and FGD guides were developed in English in accordance with previous FB interviews and the objectives of this study. The interview guides were reviewed for linguistic and cultural appropriateness. The guides were translated into Vietnamese, discussed, and revised by the research team, who had knowledge and experience on the study topic and participants. The guides were piloted among 5 patients who met the study inclusion criteria but received treatment from another MMT clinic and did not participate in the study. The IDIs and FGDs were facilitated in Vietnamese by 2 interviewers who were trained in qualitative research methods and had extensive experience working with people living with HIV who inject drugs. Each IDI had an average duration of 50 (SD 18) minutes. The average FGD duration was 1.5 (SD 25) hours. All IDIs and FGDs were conducted in Vietnamese, audio recorded, and transcribed verbatim. The interview and FGD transcriptions were deidentified and translated into English. The analysis was conducted according to the applied thematic analysis approach [37]. Thematic analysis was performed using NVivo (version 12.0; QSR International). All qualitative research was guided by the consolidated criteria for reporting qualitative studies, a 32-item checklist [38]. In all, 4 study members with experience in qualitative data analysis were involved in the data analysis process. The list of parent and child codes was developed and agreed upon by the coders. Parent and child codes and definitions were created a priori and were derived from the interview guide and literature review. Furthermore, 20% of the transcripts were double coded. Intercoder reliability was assessed after all coders coded the first 2 transcripts. A total of 2 meetings were held among the coders to discuss the coding results item by item. The official coding process began after all 4 coders agreed on the coding method and procedure. The codebook was refined throughout the analysis process and the findings were presented, discussed, and agreed upon by the study team.

**Phase 4: Production**

In all, 3 external content experts reviewed the phase 3 results. The original FB manual study procedures and content were modified according to expert comments. On the basis of the content experts’ review, the manual was tailored for use by both health care providers and lay counselors. In addition, the experts’ feedback informed the study team to include relevant matters regarding injecting drugs and HIV transmission in Vietnam. Finally, all the case studies were revised to make them suitable for participants in Vietnam.

**Phases 5 and 6: Topical Experts and Integration**

We solicited and integrated input and feedback on the first draft of the adapted FB program from 2 investigators, 2 creators of FB, and 1 topical expert on behavioral programs for people living with HIV and SUD from Hanoi Medical University (HMU). This feedback informed the development of the next draft of the FB manual. Culturally, Vietnamese illustrations were designed to replace Zimbabwean illustrations and were added to the second manual draft in English. A total of 2 translators of the study team had experience with people living with HIV who inject drugs and mental health and were responsible for translating the second draft. The Vietnamese topical expert revised the translated draft to ensure conceptual equivalency. The team then revised the translation again and agreed on the second FB manual version.

**Phases 7 and 8: Training and Testing**

Using the second adapted version, the creators of FB led a Training of Trainers session for 6 study team members in Vietnam and 3 people from HMU via a 5-day virtual training course. Trainers in Vietnam then provided in-person 5-day training for lay counselors and health care counselors in MMT clinics. The second draft was then piloted with 5 eligible patients to further refine the session content. The process culminated in a final version of the adapted FB manual that would be ready for formal evaluation in an RCT study (Table 1).

<table>
<thead>
<tr>
<th>Phase</th>
<th>Method</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Assessment of new study population</td>
<td>Assessed needs of people living with HIV on MMT\textsuperscript{a} for CMDs\textsuperscript{b}</td>
<td>Completed in 2019</td>
</tr>
<tr>
<td>2: Decision on choosing EIP\textsuperscript{c} for adaptation</td>
<td>Reviewed EIPs for CMDs then decided to choose FB\textsuperscript{d} counseling CMDs for community people of Zimbabwe to adapt for counseling CMDs for people living with HIV on MMT</td>
<td>Completed in 2019</td>
</tr>
<tr>
<td>3: Adaptation</td>
<td>Conducted IDI\textsuperscript{e} with 12 people living with HIV on MMT, 5 family members and caretakers, 4 clinic directors, and 2 FGDs\textsuperscript{f} with MMT health care providers (each group had 5 people); analyzed results by themes and reported results</td>
<td>Original manual</td>
</tr>
<tr>
<td>4: Production</td>
<td>Presented results of phase 3 to 2 creators of FB and 2 investigators and 1 expert in HMU\textsuperscript{g}; recorded and reported comments of 2 investigators and topical experts for production; reviewed original FB manual to modify per results above and received investigators’ agreement on adaptation into the first draft manual</td>
<td>First draft</td>
</tr>
<tr>
<td>5: Topical experts</td>
<td>Sent the first draft manual to 3 topical experts and 2 investigators for comments and revision</td>
<td>First draft</td>
</tr>
<tr>
<td>6: Integration</td>
<td>Integrated experts’ and investigators’ comments to make the second draft, added Vietnamese illustrations to make the third draft in English, and sent the third draft to experts and investigators again for comments; translated the third draft to Vietnamese; sent the Vietnamese version of the third draft to a Vietnamese topical expert for revision and to ensure conceptual equivalency</td>
<td>Second draft</td>
</tr>
<tr>
<td>7: Training</td>
<td>Creators of FB trained for trainers in Vietnam; trainers trained for counselors</td>
<td>Third draft</td>
</tr>
<tr>
<td>8: Testing</td>
<td>Piloted the third draft for 5 people living with HIVs on MMT and having CMDs; revised third draft to make final manual</td>
<td>Final</td>
</tr>
</tbody>
</table>

\textsuperscript{a}MMT: methadone maintenance treatment.  
\textsuperscript{b}CMD: common mental disorder.  
\textsuperscript{c}EIP: evidence-informed program.  
\textsuperscript{d}FB: Friendship Bench.  
\textsuperscript{e}IDI: in-depth interview.  
\textsuperscript{f}FGD: focus group discussion.  
\textsuperscript{g}HMU: Hanoi Medical University.

Ethics Approval
The study protocol is available at ClinicalTrials.gov (NCT04790201). The study protocol, interview guides, and informed consent forms were approved by the institutional review boards at University of North Carolina at Chapel Hill on August 24, 2020 (study 20-1689), and HMU on June 19, 2020 (decision 119 ĐHYHN). All the study participants provided written informed consent in Vietnamese.

Results
Phase 3: Adaptation
Overview of the Study Participants
We approached 67 patients using purposive sampling representing people living with HIV on MMT in 4 MMT clinics in Hanoi to answer the DASS-21 questions. Less than one-third of the patients approached (19/67, 28%) met the DASS-21 threshold CMD scores. Of 19 eligible patients, 12 (63%) agreed to be interviewed. A patient was female. The average age of the patients was 44 years. In all, 83% (10/12) of the patients had symptoms of depression, 83% (10/12) had symptoms of anxiety, and 33% (4/12) had symptoms of stress at screening (Table 2). The age of family members and caretakers ranged from 60 to 73 years. All participants were either retired or unemployed. Health care providers included 2 MMT clinic directors, 1 associate director of CDC Hanoi, and 1 deputy director of the Hanoi Department of Health. A total of 10 health care workers, including MMT physicians and nurses in MMT clinics aged 35 to 66 years, participated in 2 FGDs.

The adaptation phase in our study aimed to explore the most culturally appropriate way to adapt FB in Vietnam. Using IDIs with patients, their family members or caretakers, clinic directors, and FGDs with health care providers, we addressed the who, where, when, and how regarding the administration of the FB program in the MMT clinics.
Table 2. Demographic characteristics of participants (N=12).

<table>
<thead>
<tr>
<th>Demographic characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>11 (92)</td>
</tr>
<tr>
<td>Female</td>
<td>1 (8)</td>
</tr>
<tr>
<td><strong>Age (years), mean (SD; range)</strong></td>
<td>44 (6; 35-56)</td>
</tr>
<tr>
<td><strong>Employment, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>6 (50)</td>
</tr>
<tr>
<td>Working full-time or part-time</td>
<td>6 (50)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>5 (42)</td>
</tr>
<tr>
<td>Married</td>
<td>3 (25)</td>
</tr>
<tr>
<td>Divorced or separated</td>
<td>4 (33)</td>
</tr>
<tr>
<td><strong>Years on ART</strong>, mean (SD)</td>
<td>8 (6)</td>
</tr>
<tr>
<td><strong>Years on MMT</strong>, mean (SD)</td>
<td>5 (2)</td>
</tr>
<tr>
<td><strong>DASS-21</strong>, n (%)</td>
<td></td>
</tr>
<tr>
<td>Symptom of depression</td>
<td>10 (83)</td>
</tr>
<tr>
<td>Symptom of anxiety</td>
<td>10 (83)</td>
</tr>
<tr>
<td>Symptom of stress</td>
<td>4 (33)</td>
</tr>
<tr>
<td><strong>Number of types of different CMD$^d$ symptoms, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>One symptom</td>
<td>3 (25)</td>
</tr>
<tr>
<td>Two symptoms</td>
<td>6 (50)$^e$</td>
</tr>
<tr>
<td>Three symptoms</td>
<td>3 (25)</td>
</tr>
</tbody>
</table>

$^a$ART: antiretroviral therapy.
$^b$MMT: methadone maintenance treatment.
$^c$DASS-21: Depression, Anxiety, and Stress Scale-21 items.
$^d$CMD: common medical disorder.
$^e$Overall, 5 had depression and anxiety 1 had anxiety and stress.

Who Should Deliver the Program

When being asked who (health staff person or peer) should provide the FB program for people living with HIV on MMT, the study respondents gave mixed reviews of who would be best for the FB program. Broadly, it appeared that directors and focus group respondents saw more advantages with a staff member as the program provider. Among family members, caretakers, and patient respondents, there was no distinct, overarching opinion on a preferred program provider. A total of 2 directors and 2 patient respondents shared that the professionalism of health care providers would be beneficial. A family member stressed that staff members should be purposively selected and need certain personality traits to be as effective as lay counselors, including honesty and willingness to share their experiences with patients in FB. Clinic providers noted that clinical experience with ART and MMT management are also helpful skills for a potential program provider. They shared that their mental health counseling experience may not be at the level needed for this type of program, as exemplified in focus group FGD_01:

> It's certainly good. We haven't been trained in depression. We haven't had a chance to attend any depression courses. That’s why we can’t help our patients. We just told the patient’s family to take the patient to the psychiatric clinic. Without experience, we can’t support more than that. I have worked here for many years, but there has been no training course for depression.

A director emphasized key aspects to improve and that staff members as program providers may need further training and understanding, enthusiasm, and gentleness:

> Knowledge is always along with skills. They must be trained in skills and knowledge, when they understand, they can do it. Their attitude and their speech towards the patients must be from their hearts. If it’s a real feeling, the patients will believe in the health workers. Patients are sensitive, sometimes only a glance could affect them. [Director_IDI_D03]

For the program led by peers, participants from different groups (1 director, 4 FGD participants, 2 family members, and 7
patients) agreed that peers could understand patients very well. Peers and patients had many similarities in terms of life, health, and living conditions. Peers understood the patient’s language and experienced similar things, both good and bad. In addition, formality was unnecessary between them, making it easier for patients to talk and share their opinions with peers, which could not be achieved with staff members:

I think it’s feasible because they are the persons same as us, so they can understand the mindset of sick people like me, you are not sick so you cannot understand our thoughts. [Patient_IDI_102]

Moreover, peers had a better understanding of patients’ health conditions from their perspectives and experiences. All of these built more trust between peers and patients, facilitating the counseling sessions. However, according to clinic providers and directors, to successfully implement the FB program, peers had to have a serious attitude and commitment to the task. In doing so, they need training on how to provide counseling and care for patients to fulfill the tasks of a counselor. In addition to their practical experience, peers should update their knowledge about the topic and improve communication skills to effectively transfer what they have learned to the patients. Only 1 clinic health care provider and 1 patient were concerned that peers might not be able to fully understand their patients’ conditions. Because of their addiction and medication and their mental state, it might be difficult to find enthusiastic and suitable peers to lead the program:

Because a peer can change differently day by day, it’s difficult to find an appropriate methadone peer, but it is easier to find an ART peer. In my clinic, there is a peer who also counsels other patients...But we should think carefully about selecting methadone peers, we should consider medical staff doing the programme because we don’t know if they [peers] are stable or not. We may have to change to others if we choose them. [FGD_01]

In addition, although peers might receive training to do the job, they lack practical experience in caring for patients with different types of mental problems. A dominant theme among the patient respondents was the importance of the personality traits and counseling approach of the program provider. Patient respondents stressed the need for counselors who are gentle, empathetic, and understanding:

Someone like me needs affection, gentle speech, even a slightly unappropriated attitude will lead to my rejection. I need them to be considerate and gentle. [Patient_IDI_303]

Considering the insights from interviews and FGDs in the adaptation phase, the study selected health care providers and people living with HIV or community members who were trusted by the study participants to work as study counselors in the RCT.

When to Deliver the Program

All clinic directors, health care providers, family members, and 9 participants (14 references from clinic directors, 15 references from FGDs, 10 references from family members, and 46 references from patients) reported their concerns about what time to implement the FB program. They all shared that as many patients work, it might be challenging to find a suitable time for them to attend the FB program. Most of the patients suggested that the counseling sessions should be in the morning when they come to take methadone:

Because we usually take the medicine in the morning and after that, we can do activities. In the afternoon, we can’t wait, at that time we have to go home, so the most convenient time is the morning. [Patient_IDI_301]

A patient shared that many patients could not sit for a long time, which would affect their participation in the counseling sessions. The patient believed that the counseling time should be arranged to suit the patient’s employment and examination time, such as on weekends. For clinic directors and health care providers, apart from patients’ employment, the timetable for clinic providers is very strictly scheduled; thus, time arrangement for the clinic providers to join the FB program is problematic. Providers thought that it is best to conduct FB during office hours to ensure convenience and safety for counselors and patients:

If we do it outside of office hours, we should consider the safety of people here. If something happens, we are also affected. So we should do counseling in the daytime, not nighttime because we can’t control if they do something bad or not. [FGD_01]

However, it is necessary to arrange counseling time in advance so that clinic staff do not have many overlapping tasks and have enough time for counseling patients:

So if we have such a programme then we have to have a better management of our time so that we can do it. [FGD_01]

Where to Deliver the Program

All directors and most health care providers said that the place of program should be at the MMT clinic, as this is the place MMT patients visit every day. At MMT clinics, there were examination rooms; therefore, confidentiality and privacy would be ensured during the counseling sessions. Counseling outside the MMT clinic might be noticed by community members, which patients feared would cause them to be disliked. In addition, it is illegal for drug addicts to congregate in public spaces:

It’s not possible to gather in the community. It is against the law, they are already drug addicts, if they gather that is against the law, and people in the community don’t like that, so the methadone treatment facilities are the most suitable places. In the community, there are also a few places like that, but I only see their family come, I never see them [patients/people on treatment]. So, the medical facilities are still the best place. [Director_IDI_D03]

Half of the patients (6/12, 50%) and almost all family members (4/5, 80%) preferred to receive counseling in MMT clinics for convenience and safety for both counselors and patients,
ensuring privacy for patients and having timely and relevant health services if needed. The other respondents expressed that they needed to have a private and confidential place for counseling. They also suggested that the program only needs to designate a random room in the clinic as the counseling room and that it would be great if the room could have materials or visual aids to refer to or use.

What and How to Deliver the Program

All 4 directors mentioned that the program should be facilitated by the medical leadership. According to the directors and health care providers, to have the FB program implemented, it is crucial that the program follows all the administrative procedures of the MMT clinics before starting it. It should have a clear and detailed work plan, in which who, how, where, and when to conduct each activity are specified. Importantly, it is necessary to know about the people living with HIV on MMT and the severity of their CMD symptoms to have appropriate approaches and program sessions for them:

If the programme is implemented in the clinic, in general, in terms of administrative procedures, I also said earlier that there must be a direction. As for the implementation, first, you must have purposes and goals, then it is necessary to train staff, then find the target group we need to consult, and then plan a schedule for specific activities, jobs. [Director_IDI_01]

From the perspective of patients and their family members, though they believed that patients would benefit from a program such as FB, they thought the patients would participate in the program if they received compensation for their time and effort in participating in FB sessions (3/31, 10%) and did not have to pay or have any additional restrictions for program participation (2/31, 6%):

If only he participates without any condition from the programme, that’s very good. [Family_IDI_1011]

Participants thought that not having any additional challenges related to employment (10/31, 32%), financial issues (12/31, 39%), and transportation (2/31, 6%) would facilitate participation. Patients expected the program to be new and be organized interestingly and attractively:

I can’t imagine it yet, but I want new and interesting methods. We will love it and if the programme have people like you, we will join. [Patient_IDI_101]

Family members and patients discussed the following recommendations: the program materials should be easy to understand; fruits, cake, and drinks should be provided; patients who actively participated in the program activities and adhered to treatment should be rewarded and praised; and if possible, the program could assist patients with job opportunities or friend-making.

Both clinic providers and patients agreed that they preferred to delicately, rather than directly, discuss mental health to help patients feel comfortable. A health care provider suggested the use of words and language that is familiar to patients to help them feel more comfortable joining the program:

I think their language when talking together is very important, they want to release and use their language to feel comfortable but they can’t do that when talking to medical staff. [FGD_01]

A patient even suggested excluding the name of the program if possible. None of the suggestions for the language surrounding FB included “mental illness” in the program name.

From the perspective of family members, words in Vietnamese that are simple and easy to understand with positive meaning were preferable, and sensitive words that might worsen patients’ mental state should be avoided. A health care provider and a patient thought that the name “Friendship Bench” was beautiful and acceptable, as it did not mention the problem and had a pleasant feeling:

The Friendship bench name is also very beautiful and good. [FGD_F01]

Phase 4: Production

The results of the qualitative research in phase 3 were presented in a web-based meeting with 2 investigators, 2 creators of FB, and 1 expert from HMU to reach a consensus on the necessary adaptation of the FB manual. The original FB was adapted for use by both health care providers and lay counselors in Vietnam. Table 3 describes the details of FB adaptation. Overall, the manual illustrations of counselors were changed from Zimbabwean to Vietnamese and included both sexes. The illustrations of case studies have also been redrawn. The title “Lay Health Workers” was changed to “Counselors” to suit both groups of health care counselors and lay counselors for the study in Vietnam. The training curriculum was shortened from 8 to 5 days, and the new training program was written in a web-based format. The preface cited additional information about HIV infection and substance use patterns in Vietnam and provided a brief introduction to the adaptation of the FB manual to the study in Vietnam.

We retained the content of 8 out of the 23 chapters in the original manual. The 8 chapters included PST, counseling skills, FB cards, emotions, stabilization, strong emotional reactions, men’s health-seeking behaviors, and self-care. We completely removed 6 of the 23 chapters, namely epilepsy, belief in supernatural powers, psychosis, home visits, group circle, and proverbs. We edited the contents in 8 of the 23 chapters. Key edits included using the words Common Mental Disorders to refer to a collective term of depression, anxiety, and stress-related disorders instead of “kufungisisa” (Zimbabwean for depression and anxiety), replacement of SSO–14 to DASS–21, and suicide risk assessment and management procedures, which were used in the study in Vietnam. In addition, we excluded sessions on abuse of bronchodilators, alcohol and pregnancy, and information about antipsychotics, as they were not related to the study population. Information about the use of FB on tablets was removed, as the Vietnam FB program used paper forms to record counseling session information. Information on HIV transmission through injection, substance abuse disorders caused by injecting drugs, and amphetamine-type stimulants, which were common issues of people who inject drugs—the study population in the RCT—was added.

Table 3 describes the details of FB adaptation. Overall, the manual illustrations of counselors were changed from Zimbabwean to Vietnamese and included both sexes. The illustrations of case studies have also been redrawn. The title “Lay Health Workers” was changed to “Counselors” to suit both groups of health care counselors and lay counselors for the study in Vietnam. The training curriculum was shortened from 8 to 5 days, and the new training program was written in a web-based format. The preface cited additional information about HIV infection and substance use patterns in Vietnam and provided a brief introduction to the adaptation of the FB manual to the study in Vietnam.

We retained the content of 8 out of the 23 chapters in the original manual. The 8 chapters included PST, counseling skills, FB cards, emotions, stabilization, strong emotional reactions, men’s health-seeking behaviors, and self-care. We completely removed 6 of the 23 chapters, namely epilepsy, belief in supernatural powers, psychosis, home visits, group circle, and proverbs. We edited the contents in 8 of the 23 chapters. Key edits included using the words Common Mental Disorders to refer to a collective term of depression, anxiety, and stress-related disorders instead of “kufungisisa” (Zimbabwean for depression and anxiety), replacement of SSO–14 to DASS–21, and suicide risk assessment and management procedures, which were used in the study in Vietnam. In addition, we excluded sessions on abuse of bronchodilators, alcohol and pregnancy, and information about antipsychotics, as they were not related to the study population. Information about the use of FB on tablets was removed, as the Vietnam FB program used paper forms to record counseling session information. Information on HIV transmission through injection, substance abuse disorders caused by injecting drugs, and amphetamine-type stimulants, which were common issues of people who inject drugs—the study population in the RCT—was added.

Table 3 describes the details of FB adaptation. Overall, the manual illustrations of counselors were changed from Zimbabwean to Vietnamese and included both sexes. The illustrations of case studies have also been redrawn. The title “Lay Health Workers” was changed to “Counselors” to suit both groups of health care counselors and lay counselors for the study in Vietnam. The training curriculum was shortened from 8 to 5 days, and the new training program was written in a web-based format. The preface cited additional information about HIV infection and substance use patterns in Vietnam and provided a brief introduction to the adaptation of the FB manual to the study in Vietnam.

We retained the content of 8 out of the 23 chapters in the original manual. The 8 chapters included PST, counseling skills, FB cards, emotions, stabilization, strong emotional reactions, men’s health-seeking behaviors, and self-care. We completely removed 6 of the 23 chapters, namely epilepsy, belief in supernatural powers, psychosis, home visits, group circle, and proverbs. We edited the contents in 8 of the 23 chapters. Key edits included using the words Common Mental Disorders to refer to a collective term of depression, anxiety, and stress-related disorders instead of “kufungisisa” (Zimbabwean for depression and anxiety), replacement of SSO–14 to DASS–21, and suicide risk assessment and management procedures, which were used in the study in Vietnam. In addition, we excluded sessions on abuse of bronchodilators, alcohol and pregnancy, and information about antipsychotics, as they were not related to the study population. Information about the use of FB on tablets was removed, as the Vietnam FB program used paper forms to record counseling session information. Information on HIV transmission through injection, substance abuse disorders caused by injecting drugs, and amphetamine-type stimulants, which were common issues of people who inject drugs—the study population in the RCT—was added.
Table 3. Summary of changes in Friendship Bench (FB) manual.

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Name of chapter</th>
<th>Revision</th>
</tr>
</thead>
</table>
| N/A     | Cover page     | • Changed the illustration from Zimbabwean to Vietnamese  
          |                | • Changed the word “Lay health workers” to “Counselors” to suit both groups of counselors of the study in Vietnam |
| N/A     | Introduction   | • Rewrote the preface to fit the purpose of the program in Vietnam |
| 1       | Psychoeducation| • Used the words “Common Mental Disorder” to refer to a collective term of depression, anxiety, and stress-related disorders instead of “kufungisisa” (Zimbabwean for depression and anxiety).  
          |                | • Removed information about tablets |
| 2       | Core competencies| • Changed chapter title from “Lay health workers” to “Counselors”  
          |                | • Replaced the Zimbabwean CMD\(^b\) screening SSQ\(^c\) with DASS-21\(^d\) for research in Vietnam |
| 3       | Mental illness or mental neurological substance use disorders | • Removed information about postpartum depression, fatherhood, sexual partners, and family |
| 4       | Epilepsy       | • Removed completely |
| 5       | Medication     | • Kept information about ART\(^e\)  
          |                | • Added information about MMT\(^f\) |
| 6       | Belief in supernatural powers | • Removed completely |
| 7       | HIV and mental health | • Added information on HIV transmission through injection  
          |                | • Removed information about misunderstandings about church teachings |
| 8       | Substance use disorders | • Added information on substance use disorders caused by injecting drugs  
          |                | • Added information about amphetamine-type stimulants  
          |                | • Removed the substance use chemicals that cause bronchodilators  
          |                | • Removed alcohol and pregnancy |
| 9       | Psychosis      | • Removed completely |
| 10      | Problem-solving therapy | • Retained |
| 11      | Questionnaires | • Replaced Screening CMD by SSQ in Zimbabwe by DASS-21 in study |
| 12      | Counseling skills | • Retained |
| 13      | FB cards       | • Retained |
| 14      | Emotions       | • Retained |
| 15      | Stabilization  | • Retained |
| 16      | Strong emotional reactions | • Retained |
| 17      | Men’s health-seeking behaviors | • Retained |
| 18      | Suicide assessment | • Replaced suicide assessment of SSQ by the suicide assessment of the study in Vietnam |
| 19      | Supervision    | • Changed monitor titles to the study staffs in Vietnam |
| 20      | Home visits    | • Removed completely |
| 21      | Group circle   | • Removed completely |
| 22      | Self-care      | • Retained |
| 23      | Proverbs       | • Removed completely |
All cases were revised, and the PST case was rewritten. In the original FB, the PST case featured a common problem in the Zimbabwe community, and the main character was a person with an unknown HIV status, a person with SUD, or a woman experiencing domestic violence. The new character in the adaptation was people living with HIV who inject drugs with short-term, solvable problems about employment, family matters, finances, and relationships with neighbors.

Phase 5: Topical Experts

The first English version was sent to 2 investigators, 2 creators of FB, and 1 expert from HMU for viewing and editing. All of them are associate professors with intensive expertise in the areas of mental health, HIV or AIDS, and substance use. They have led clinical studies and trials on a global scale. Experts advised correcting cases of PST to adjust them to Vietnamese culture. They recommended not changing the in-person counseling procedures and having up to 6 individual counseling sessions for each patient. An FB creator wrote scripts for the counselors to be included in the draft. The creator also advised maintaining the PST structure, with 4 main parts in each counseling session. The 4 main parts were the following: open your mind, make a list of problems, plan to solve the problem according to the Specific, Measurable, Achievable, Realistic, and Timely method, and encourage the clients to implement the plan. The 4 main parts of the PST were implemented into 7 small steps: (1) How does the client deal with problems? (2) How to recognize a problem? (3) How to select a problem, find the goal, and define the problem? (4) How to brainstorm for solutions? (5) How to select a solution? (6) How to make a Specific, Measurable, Achievable, Realistic, and Timely action plan? and (7) Did it work?

Phase 6: Integration

On October 20, 2021, a meeting was held on the web with 3 content experts, the study team in Vietnam, and 2 investigators. The investigators agreed on continuing to revise the case example of a typical people who inject drugs who encounters common mental problems in their life. These problems should be specific, simple, and solvable within 1 week. For example, “I want to quit using drugs for the next 1 year because next week is my son’s birthday.” The study team agreed to remove the group discussion portion, as it would be ineffective where MMT clinics are far from each other, without finding additional study resources. The study team then synthesized the expert revision to create the second draft in English. Illustrations of Vietnamese people were added to the second English version to make the third version which was sent to 3 content experts and 2 investigators to review and finalize. The third English version was translated into Vietnamese with the same content and images. A Vietnamese topical expert reviewed the translated third version to ensure conceptual equivalency.

Discussion

Principal Findings

This study successfully illustrates the process of operationalizing the ADAPT-ITT framework to adapt to a mental health program in Vietnam. The outcome of this study was an adapted FB manual to be used for training and piloting to create the final program manual.

We note that our study departs from the standard ADAPT-ITT framework by selecting an EIP that does not come from the CDC database of EIPs [26] and by adapting a program addressing CMDs rather than HIV prevention per the usual ADAPT-ITT goals [25]. Nevertheless, our study was methodologically aligned with the ADAPT-ITT goals, as we chose a proven, effective program for adaptation. FB improved CMDs among people living with HIV in Zimbabwe and Malawi, and we hope to use FB to address CMDs among people living with HIV on MMT in Vietnam.

The application of the ADAPT-ITT framework for the systematic adaptation of EIP can vary depending on the program, population, and resources [28,30,39]. A key strength of this study is that it used a linear approach to adapt the ADAPT-ITT framework on a small scale [28]. Results of the previous phases provided informative contributions for the next phase [40] in terms of the program procedures and content of the program manual. Other study and adaptation strengths were the translation into Vietnamese and the creation of local
illustrations to increase the clarity of the concepts while maintaining the original meaning of the English version.

The results from phase 3 informed the program procedures built upon the existing resources and administrative strength of MMT clinics [40]. As a result of adaptation, the study will choose health care providers and people living with HIV or community members who are trusted by the study participants to work as study counselors. The counselors will be trained and managed closely to ensure the quality of counseling. The counseling locations must be private and safe in MMT clinics. Counseling appointments will be scheduled between counselors and patients in parallel. Standardized criteria of counselors to deliver the program are defined as being understanding and having the trust of the patients but not necessarily having a high level of counseling techniques before training. Instead of using tools and procedures in the original FB program, the study will use questionnaires and protocols relevant to the study participants.

Regarding the content of the manual, the ADAPT-ITT framework guided the appropriate adaptation of the program while maintaining the core components of the PST of the original program throughout the counseling techniques in all program sessions [25,26]. The strength of our adaptation process was that the PST of the original FB fits the goals of reducing CMDs for people living with HIV on MMT in Vietnam; therefore, we did not have to change PST counseling as well as other chapters mentioning CMDs, HIV, and SUD in the original program that address common problems that global and Vietnamese people living with HIV and people who inject drugs encounter [41,42]. Feedback from topical experts informed the study team to alter some contents of the original program to ensure that it was suitable for Vietnamese people living with HIV on MMT and in MMT settings [27,40]. Specifically, unlike the original program conducted in Zimbabwe, we cut content that is not common in Vietnamese culture such as a church, beliefs in supernatural power, and praying together. The study also changed the illustrations from Zimbabwean to Vietnamese. Information from a qualitative study in phase 3 about common health problems faced by people living with HIV on MMT (the results are not reported within the scope of this paper) helped the study team rewrite sample PST cases to reflect typical problems in the key population. In addition, because of the new study with people living with HIV on MMT in Vietnam, we added additional guidelines on HIV, injecting drugs, and SUD to the original program to make it more culturally relevant and relevant to the study population. As recommended by topical experts, we dropped group activities that were proven more adaptive to the Zimbabwean setting and chapters less relevant to the target population (fatherhood, alcohol and pregnancy, epilepsy, and psychosis).

Limitations
This study has some limitations, the first being that the process of adaptation and production has not yet tested the feasibility and acceptability of the adapted manual. In addition, the semistructured interviews had not presented the original manual to the study participants. Therefore, we did not have comments from them on the content of the manual. Instead, the content of the manual was reviewed and revised in detail according to the study team’s revisions and experts’ comments. There might be challenges such as following the structure of FB counseling sessions, applying relevant counseling skills required in FB, following PST methods for health care, and lay counselors using the adapted manual as a resource material in conducting PST counseling sessions as they first work with this approach. As such, the adapted manual will be used to train counselors and piloted with the study participants of the future RCT to receive feedback to finalize the manual. Study participants and counselors can provide their feedback during the pilot phase. Future RCTs will test the acceptability and fidelity of the finalized adapted program.

Conclusions
This initial exploratory study demonstrates a successful process of following the ADAPT-ITT to adapt a proven mental health program for people living with HIV on MMT in Vietnam. This study selected and culturally adapted an evidence-informed PST program to improve CMDs among people living with HIV on MMT in Vietnam. The adaptation of the program through qualitative interviews and discussions with stakeholders and study participants and the use of feedback to tailor the program procedures allowed us to identify and preemptively address potential barriers to implementation. The production, integration, and expert input phases were used to tailor the manual to reflect typical manageable problems that people living with HIV on MMT encounter daily. If the adapted FB manual is acceptable and feasible, it may be used in MMT clinics in Vietnam to reduce CMDs for patients on MMT, which would contribute to the effectiveness of drug treatment and ART.

Acknowledgments
The authors are grateful for the support of the leaders of the Hanoi Department of Health and the Centers of Disease Control Hanoi and the 4 methadone maintenance treatment clinics in Hanoi for the implementation of this study. The authors would like to thank interviewers Nguyen Nhu Trang and Vu Thi Thanh Nhan who conducted the interviews for this study. Finally, the authors would like to thank the participants for their invaluable contribution to this research.

Authors’ Contributions
BNG and BWP obtained funding and designed the study. HVT and HTTN oversaw training and data collection. HVT collected the data. TTTT, HTTN, TRF, and KRL coded and analyzed all data. KRL created the analysis plan, reviewed the coding, and reviewed memos to measure consistency and reliability. HVT drafted the manuscript. BNG, BWP, KRL, TRF, VFG, RV, HTH,
Conflicts of Interest
None declared.

References


**Abbreviations**

ADAPT-ITT: Assessment-Decision-Adaptation-Production-Topical Experts-Integration-Training-Testing
ART: antiretroviral therapy
CDC: Centers for Disease Control and Prevention
CMD: common mental disorder
DASS-21: Depression, Anxiety, and Stress Scale-21 items
EIP: evidence-informed program
FB: Friendship Bench
FGD: focus group discussion
HMU: Hanoi Medical University
IDI: in-depth interview
MMT: methadone maintenance treatment
PST: problem-solving therapy
RCT: randomized controlled trial
SSQ-14: Shona Symptoms Questionnaire
SUD: substance use disorders

©Ha V Tran, Ha T T Nong, Thuy T T Tran, Teresa R Filipowicz, Kelsey R Landrum, Brian W Pence, Giang M Le, Minh X Nguyen, Dixon Chibanda, Ruth Verhey, Vivian F Go, Hien T Ho, Bradley N Gaynes. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.07.2022. 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.
Preliminary Real-World Evidence Supporting the Efficacy of a Remote Neurofeedback System in Improving Mental Health: Retrospective Single-Group Pretest-Posttest Study

Jocelyne C Whitehead¹,², PhD; Ron Neeman², MA; Glen M Doniger², PhD

¹Integrated Program in Neuroscience, McGill University, Montreal, QC, Canada
²Myndlift Ltd, Tel Aviv, Israel

Corresponding Author:
Jocelyne C Whitehead, PhD
Integrated Program in Neuroscience
McGill University
845 Rue Sherbrooke Ouest
Montreal, QC, H3A 0G4
Canada
Phone: 1 866 994 3210
Email: jocelyne.whitehead@mail.mcgill.ca

Abstract

Background: Neurofeedback training (NFT) has been shown to be effective in treating several disorders (e.g., attention-deficit/hyperactivity disorder [ADHD], anxiety, and depression); however, little is currently known regarding the effectiveness of remote NFT systems.

Objective: This retrospective study provides real-world data (N=593) to assess the efficacy of app-based remote NFT in improving brain health and cognitive performance.

Methods: Improvement was measured from pre- to postintervention of in-app assessments that included validated symptom questionnaires (the 12-item General Health Questionnaire, the ADHD Rating Scale IV, the Adult ADHD Self-Report Scale, the 7-item Generalized Anxiety Disorder scale, and the 9-item Patient Health Questionnaire), a cognitive test of attention and executive functioning (i.e., continuous performance task), and resting electroencephalography (EEG) markers. Clinically significant improvement was evaluated using standard approaches.

Results: The greatest improvement was reported for the anxiety questionnaire, for which 69% (68/99) of participants moved from abnormal to healthy score ranges. Overall, adult and child participants who engaged in neurofeedback to improve attention and executive functions demonstrated improved ADHD scores and enhanced performance on a cognitive (i.e., response inhibition) task. Adults with ADHD additionally demonstrated elevated delta/alpha and theta/alpha ratios at baseline and a reduction in the delta/alpha ratio indicator following neurofeedback.

Conclusions: Preliminary findings suggest the efficacy of app-based remote neurofeedback in improving mental health, given the reduced symptom severity from pre- to postassessment for general psychological health, ADHD, anxiety, and depression, as well as adjusted resting EEG neural markers for individuals with symptoms of ADHD. Collectively, this supports the utility of the in-app assessment in monitoring behavioral and neural indices of mental health.

(JMIR Form Res 2022;6(7):e35636) doi:10.2196/35636

KEYWORDS
EEG biofeedback; remote care; neurofeedback; attention-deficit/hyperactivity disorder; delta/alpha ratio

Introduction

Background

Neurofeedback training (NFT) is considered a primary or supplementary treatment for a number of disorders, including attention-deficit/hyperactivity disorder (ADHD) [1-5], anxiety [6-9], and depression [7,8,10]. The American Academy of Pediatrics [11] provided a “level 1 best support” rating of NFT as a safe and effective evidence-based therapy for childhood ADHD. Nonetheless, several significant barriers prevent patients from receiving quality neurofeedback therapeutics; for example,
electroencephalography (EEG) systems are expensive, complex, and often only accessible at health care clinics. A recent pilot study [12] provided encouraging evidence for the efficacy of therapist-guided NFT, suitable for remote home-based use. Findings showed improved ADHD symptomatology in a small cohort of children after 9 weeks of NFT. The system was designed as an affordable convenient wireless alternative to clinic-based EEG. NFT users regulate neural activity through operant conditioning, which can lead to morphological changes in the brain [13,14] and calmer, more focused cognitive, affective, and physical functioning. Currently, little is known about the effectiveness of NFT systems in the field [15]; therefore, this retrospective open-label pilot study offers real-world data supporting the efficacy of remote NFT in improving brain health.

Mental Health Improvement in Real-World Settings

Unlike standard EEG systems, Myndlift is an easy-to-use tool for patients and clinicians (Figure 1). While wearing the validated EEG headband (Muse; Interaxon [16,17]) containing four dry recording electrodes (ie, anterior frontal [AF] 7, AF8, temporal pole [TP] 9, and TP10), one ground electrode, and one auxiliary wet electrode, the patient trains with an Android or iOS app linked to the headset by Bluetooth, which delivers visual and auditory feedback during YouTube videos or specialized games. When patients’ brain waves are in the desired range, positive feedback is delivered. A therapist can set or adjust the training protocol and monitor progress remotely via a cloud-based web service. The device incorporates an app-based assessment, lasting approximately 40 minutes, completed prior to NFT (ie, baseline) and periodically over the intervention period for longitudinal tracking of improvement.

Figure 1. Myndlift in-app assessment screens. From left to right: introduction, symptom questionnaires, resting electroencephalography assessment, and cognitive task.

Real-world studies provide external validity and accurately represent the heterogeneity of a patient population [18]. From the app, real-world data were collected from more than 500 participants on outcome measures, including pre- and postintervention assessments of validated symptom questionnaires, a cognitive test of attention and executive functioning (ie, continuous performance task [CPT]), and resting EEG markers. An efficacious system could serve as a reliable, cost-effective solution for users. In-clinic NFT costs approximately US $150 to $200 per session, with a minimum of 30 to 40 sessions typically recommended. In contrast to a cost-per-session model, remote NFT could offer monthly charges, ranging from US $200 to $500.

EEG Neuromarkers of ADHD

Given the success of neurofeedback for child ADHD, more adults with ADHD are turning to NFT for treatment. Currently, 6.76% of adults worldwide—translating to 366.3 million people—are affected [19]. ADHD is commonly recognized as a hypoaroused brain state [20]. In recent years, EEG measures have provided supporting evidence for popular theoretical models of hypoactivation [21] related to core symptoms of hyperactivity, inattention, and impulsivity [22]. The hypoarousal state is best localized to frontal and posterior regions [23] (ie, neuroanatomical structures subserving attentional networks [24,25]). EEG patterns of ADHD in children are characterized by elevated low-frequency power (ie, primarily theta) and reduced relative high-frequency power (ie, alpha and beta) [23,26-28], or an elevated ratio of the two (ie, low to high frequency). The theta/beta ratio (TBR) is the most common form of NFT in treating ADHD [29,30]; however, inconsistencies in the literature suggest that TBR [31,32] may not be reliable as a diagnostic measure [33]. This may reflect EEG heterogeneity across ADHD-diagnosed individuals (eg, Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition [DSM-5] subtypes; psychiatric comorbidities; age; and sex) [34,35]. For instance, although theta and beta power differences are evident in child ADHD [31,36], a recent review [37] suggested that theta and alpha frequencies may be more reliable markers for adults. Notably, most adult studies
emphasize group differences in alpha power during eyes-closed conditions [38-43], while more recent work has identified elevated theta and delta power in adults with ADHD [44,45]. Given this evidence, this study investigates whether TBR versus the delta/alpha ratio (DAR) or the theta/alpha ratio (TAR) are biomarkers for adult ADHD. Overall, the study evaluates evidence for improvement in mental health via symptom questionnaires, a CPT, and hypothesized EEG markers. Findings have implications for the benefits of NFT and efficacy of a remote home-based system.

**Methods**

**Participants**

Participants, 13 years of age or older, signed up through their clinician or a clinician suggested by Myndlift and completed NFT at home or in clinic in a clinical care context. Informed consent was provided through the app, allowing participants' anonymized data to be used for research. Data were included for analysis if baseline (ie, preintervention) assessment was conducted after 7 or fewer NFT sessions (ie, attributed to in-app NFT tutorial). For analyses of improvement, postintervention sessions occurred 30 to 180 days after baseline with 20 or more NFT sessions completed [46]. An average of 1 or more NFT sessions per week was required for inclusion, given that effective neurofeedback requires consistency [47,48], irrespective of the neurofeedback protocol used. Data were collected via the app.

**Ethical Considerations**

Procedures were reviewed by an independent Institutional Review Board (IRB)—Pearl IRB—who permitted IRB exemption for analyses of data previously collected and deidentified, following the guidelines of the Declaration of Helsinki.

**Neurofeedback Protocol**

Participants performed neurofeedback protocols (Multimedia Appendix 1) that were customized by their clinicians and consistent with current literature [49].

**Procedure and Outcome Measures**

**Symptom Questionnaires**

The in-app assessment includes 14 brief standardized questionnaires commonly used to screen for mental health conditions. In this study, data were reported for the following five questionnaires completed at baseline and follow-up by at least 25 participants: the 12-item General Health Questionnaire (GHQ-12) [50], the ADHD Rating Scale IV (ADHD-RS-IV) [51], the Adult ADHD Self-Report Scale (ASRS) for DSM-5 [52], the 7-item Generalized Anxiety Disorder scale (GAD-7) [53], and the 9-item Patient Health Questionnaire (PHQ-9) [54]. For each questionnaire, participants filled out self-report measures based on frequency of symptom occurrence using a 4- or 5-point Likert-style scale. Total scores were calculated for use in improvement analyses. Participants engaging in neurofeedback for ADHD completed the ASRS [52] if they were 18 years of age or older; otherwise, they completed the ADHD-RS-IV. The GHQ-12, GAD-7, and PHQ-9 were completed by participants of all ages [55-58].

**Continuous Performance Task**

The assessment contained an 8-minute CPT, a behavioral test of response inhibition, in which participants are instructed to tap the screen when the target object (ie, an arrow-like shape pointing upward) is shown, but not when other stimuli appear. The interstimulus interval and presence of audiovisual distracter stimuli were varied throughout the task. Outcomes included average response time (RT) and response time variability (ie, the SD of RT [SDRT]), as well as omission and commission errors related to inattention and impulsivity, respectively [59]. This type of test is commonly used as an objective measure of attention and executive function [60-62] and has become a standard assessment tool for attentional difficulties [59,63,64].

**Resting EEG**

Resting EEG was recorded from 9 electrodes (ie, AF7, AF8, TP9, TP10, central [C] zero [z], frontal [F] z, F3, F4, and occipital [O] 1). The EEG assessment was divided into five sequential (ie, “sensing”) phases; in each phase, the auxiliary electrode was placed at a different scalp location: central (Cz), frontal (Fz), left (F3), right (F4), and back or posterior (O1). Each phase was split into eyes-closed and eyes-open blocks. A block continued until 30 seconds of clean EEG—sampled at 256 Hz—had been recorded, which typically took up to 45 seconds.

**Statistical Analysis**

**Symptom Questionnaires**

Questionnaire results were analyzed in terms of improvement in total score from pre- to postintervention, including mean change in points, effect size (ie, Cohen $d$), and percent of users with clinically significant improvement, defined as 20% improvement [65,66]. Results are presented separately for participants scoring in healthy and abnormal ranges at baseline, as per conventional clinical cutoff values. The percent of participants who shifted from abnormal to normal (ie, healthy) ranges after the intervention is also reported. Paired-samples $t$ tests (2-tailed) evaluated statistically significant improvement for each clinical measure ($P<.05$). By convention, small, medium, and large effects correspond to $d=0.2$, $d=0.5$, and $d=0.8$, respectively. For symptom questionnaires, CPT, and resting EEG analyses, multiple comparisons were corrected using the Benjamini-Hochberg (BH) method [67] to maintain a family-wise error at $P=.05$, reported as BH-adjusted $P$ values ($P_{BH}$). The Levene test assessed assumptions of equality of variance and corrected for inhomogeneities.

**Continuous Performance Task**

CPT results were analyzed for participants who completed child (ie, ADHD-RS-IV) or adult (ie, ASRS) ADHD questionnaires. Results are given in terms of improvement in RT and SDRT for correct responses (ie, shorter and less variable response times, respectively), omission errors, and commission errors. This includes mean change, in milliseconds or errors, and effect size. RT and SDRT scores were standardized by age to minimize age effects on performance [68]. Percent of participants demonstrating clinically significant improvement was reported, defined by a reliable change index (RCI) [69] that accounts for...
practice effects [70]. Exceeding a critical value of 95% for a 1-tailed test—equivalent to 1.65 SD units on a standardized z scale—indicates a significant reliable change, similar to others [71].

**Resting EEG**

Participants who completed the adult ADHD questionnaire at baseline were split into groups with “healthy” and “abnormal” ranges of values based on their score. Only participants with clean EEG signals were included (see Multimedia Appendix 2 for EEG preprocessing). Results were reported in terms of EEG amplitude (ie, Hz; relative power) for TAR, DAR, and TBR at baseline. Independent-samples t tests were conducted for each power ratio across groups (ie, healthy and abnormal values). Frequency bands were defined as follows: delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), and beta (13-30 Hz). These were averaged across frontal electrodes (ie, F3 and F4, based on the frontal nodes of the frontoparietal network [25,72] and the prevalence of a clean EEG signal) during the eyes-closed condition. Improvement analyses were conducted separately for each group and included the mean change in ratio amplitude from pre- to postintervention and associated effect size; paired-samples t tests were used to evaluate within-group changes.

**Results**

**Sample Characteristics**

Data from 560 participants met the criteria for inclusion in the analysis. Depending on clinical considerations determined by their therapist, subsets of participants completed each symptom questionnaire, CPT, resting EEG, or any combination of the three. Table 1 gives sample characteristics for each assessment component, including the NFT protocols completed by 50% or more of each sample population (Multimedia Appendix 1).

### Table 1. Sample characteristics as separated by each outcome measure and analysis.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Age (years)</th>
<th>Gender, n (%)</th>
<th>Test setting, n (%)</th>
<th>NFT protocols used in ≥50% of sample</th>
<th>No. of sessions, mean (SD)</th>
<th>Treatment duration (days), mean (SD)</th>
<th>Frequency (sessions/wk), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptom questionnaire pre-post (n=301)</td>
<td>Mean (SD)</td>
<td>Range</td>
<td>Female</td>
<td>Male</td>
<td>Clinic</td>
<td>Home</td>
<td></td>
</tr>
</tbody>
</table>
| CPT 
(ADHD; n=203)                                    | 38 (14.5)   | 13-71         | 157 (52.7)         | 141 (47.3)                          | 220 (73.1)                 | 81 (26.9)                           |                                   |
| Resting EEG baseline (adult ADHD; n=271f)          | 38 (10.9)   | 18-70         | 94 (35.2)          | 173 (64.8)                          | 87 (32.1)                  | 184 (67.9)                          |                                   |
| Resting EEG pre-post (adult ADHD; n=41f)           | 36 (9.3)    | 19-55         | 17 (42.5)          | 23 (57.5)                           | 5 (12.2)                   | 36 (87.8)                           |                                   |

aNFT: neurofeedback training.

bNo gender identity was reported by 3 participants (n=298).

cCPT: continuous performance task.

dADHD: attention-deficit/hyperactivity disorder.

No gender identity was reported by 4 participants (n=267).

EEG: electroencephalography.

N/A: not applicable; intervention details were not reported, as only preintervention values were of interest for baseline analyses.

Symptom Questionnaires

Results for participants who completed symptom questionnaires (n=301) were separated into groups with abnormal and healthy scores (Table 2). Most participants engaged in NFT protocols to reduce theta (227/301, 75.4%) and enhance high beta (248/301, 82.4%), while many who completed the PHQ-9 (76/134, 56.7%) and the ASRS (59/112, 52.7%) also performed enhanced alpha, whereas children who completed the ADHD-RS-IV also often included enhanced low beta (21/27, 78%) and enhanced sensorimotor rhythm (SMR; 16/27, 59%).

In the groups with abnormal results, all questionnaires had large effect sizes (d≥0.99 to 2.41), while the effect sizes for groups with healthy results were large only for child and adult ADHD questionnaires. Improvement in the groups with abnormal results was statistically significant for all questionnaires, with the majority (30/56, 54% to 7/77, 100%) of users demonstrating clinically significant change (ie, ≥20%) [65,66]. The most prominent improvement was observed in participants with abnormal baseline anxiety or child ADHD scores. Nevertheless,
ADHD-RS-IV findings are considered preliminary given the small sample size. Most participants (30/56, 54% to 7/7, 100%) in the groups with abnormal results shifted their values to healthy ranges at postintervention. Improvement of healthy participants was statistically significant for all questionnaires, with the majority (30/66, 45% to 14/20, 70%) demonstrating clinically significant change.

Table 2. Improvement in self-reported subjective symptoms after ≥30 days of Myndlift neurofeedback for users that scored in the healthy range, and separately for those that scored in the abnormal range (per conventional clinical cutoffs) at baseline.

<table>
<thead>
<tr>
<th>Questionnaire and group at baseline (cutoff value)</th>
<th>No. of sessions, mean (SD)</th>
<th>Treatment duration (days), mean (SD)</th>
<th>Change (points decreased), mean (SD)</th>
<th>Change T value</th>
<th>Change P value</th>
<th>Effect size, d</th>
<th>Users improved by ≥20%, n (%)</th>
<th>Abnormal to healthy results, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12-item General Health Questionnaire (maximum score = 36)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal (≥12; n=197)</td>
<td>53 (39.1)</td>
<td>94 (42.2)</td>
<td>7.8 (7.80)</td>
<td>13.94</td>
<td>&lt;.001</td>
<td>0.99</td>
<td>139 (71)</td>
<td>113 (57)</td>
</tr>
<tr>
<td>Healthy (&lt;12; n=66)</td>
<td>52 (34.3)</td>
<td>84 (36.6)</td>
<td>1.0 (4.28)</td>
<td>1.90</td>
<td>.06</td>
<td>0.23</td>
<td>30 (45)</td>
<td>N/A</td>
</tr>
<tr>
<td>ADHD Rating Scale IV (for children; maximum score = 54): preliminary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal (&gt;36; n=7)</td>
<td>49 (19.7)</td>
<td>75 (32.8)</td>
<td>19.3 (7.99)</td>
<td>6.38</td>
<td>&lt;.001</td>
<td>2.41</td>
<td>7 (100)</td>
<td>7 (100)</td>
</tr>
<tr>
<td>Healthy (≤36; n=20)</td>
<td>53 (23.2)</td>
<td>102 (37.9)</td>
<td>7.9 (8.10)</td>
<td>4.36</td>
<td>&lt;.001</td>
<td>0.98</td>
<td>14 (70)</td>
<td>N/A</td>
</tr>
<tr>
<td>Adult ADHD Self-Report Scale (maximum score =24)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal (≥14; n=56)</td>
<td>48 (25.6)</td>
<td>86 (41.5)</td>
<td>4.0 (3.81)</td>
<td>7.83</td>
<td>&lt;.001</td>
<td>1.05</td>
<td>30 (54)</td>
<td>30 (54)</td>
</tr>
<tr>
<td>Healthy (&lt;14; n=56)</td>
<td>63 (35.7)</td>
<td>97 (37.2)</td>
<td>2.1 (2.14)</td>
<td>7.38</td>
<td>&lt;.001</td>
<td>0.99</td>
<td>33 (59)</td>
<td>N/A</td>
</tr>
<tr>
<td>7-item Generalized Anxiety Disorder scale (maximum score = 21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal (≥14; n=99)</td>
<td>52 (36.7)</td>
<td>87 (40.2)</td>
<td>6.4 (5.18)</td>
<td>12.39</td>
<td>&lt;.001</td>
<td>1.24</td>
<td>82 (83)</td>
<td>68 (69)</td>
</tr>
<tr>
<td>Healthy (&lt;14; n=107)</td>
<td>55 (32.5)</td>
<td>97 (40.4)</td>
<td>1.3 (3.92)</td>
<td>3.43</td>
<td>.001</td>
<td>0.33</td>
<td>63 (59)</td>
<td>N/A</td>
</tr>
<tr>
<td>9-item Patient Health Questionnaire (maximum score = 27)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal (≥10; n=63)</td>
<td>57 (47.6)</td>
<td>88 (37.7)</td>
<td>6.2 (5.47)</td>
<td>8.94</td>
<td>&lt;.001</td>
<td>1.13</td>
<td>45 (71)</td>
<td>38 (60)</td>
</tr>
<tr>
<td>Healthy (&lt;10; n=71)</td>
<td>57 (34.7)</td>
<td>95 (39.8)</td>
<td>1.5 (4.07)</td>
<td>3.04</td>
<td>.004</td>
<td>0.36</td>
<td>49 (69)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

aReported as Benjamini-Hochberg–adjusted P values.
bN/A: not applicable; healthy subjects are already within the healthy range.
cADHD: attention-deficit/hyperactivity disorder.

Continuous Performance Task

Participants completing CPT and ADHD questionnaires performed primarily reduced theta (76/99, 77%) and enhanced high beta (81/99, 90%) protocols. Most adults also performed enhanced alpha (54/90, 60%), whereas most children also performed enhanced low beta (9/9, 100%) and enhanced SMR (7/9, 78%). Results (n=99) for average RT, SDRT, omission errors, and commission errors were divided by abnormal versus healthy scores for child and adult ADHD combined (Table 3). The greatest improvement observed, irrespective of group (ie, abnormal and healthy ADHD ranges), was in SDRT (d=1.02 and d=1.24, respectively), where nearly half of the participants (42/99, 43%) demonstrated clinically significant improvement, as indicated by the RCI. Although average RTs improved comparably (42/99, 43%), differences between pre- and postintervention were significant only for the healthy results group (d=0.56). At least one-third of users improved in their commission errors (35/99, 35%) and omission errors (45/99, 45%) from pre- to postintervention. Results from a group (n=104) with unknown ADHD assignment were comparable to those of groups with abnormal and healthy results (Multimedia Appendix 3).
Table 3. Improvement in CPT after ≥30 days of Myndlift neurofeedback (n=99) separately for healthy users that scored in the normal range for children or adults at baseline and for those in the abnormal ADHD range (per conventional clinical cutoffs).

<table>
<thead>
<tr>
<th>CPTa outcome and group results at baseline ASRSb or ADHDc-RS-IVd</th>
<th>No. of sessions, mean (SD)</th>
<th>Treatment duration (days), mean (SD)</th>
<th>Change reduc- tion, mean (SD)e</th>
<th>Change T value</th>
<th>Change P valuef</th>
<th>Effect size, d</th>
<th>Users improved (RCIf ≥1.65 SD), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average response time</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal (n=46)</td>
<td>48 (26.3)</td>
<td>85 (40.3)</td>
<td>8.9 (33.32)</td>
<td>1.80</td>
<td>.08</td>
<td>0.27</td>
<td>20 (43)</td>
</tr>
<tr>
<td>Healthy (n=53)</td>
<td>61 (31.0)</td>
<td>100 (40.0)</td>
<td>15.0 (26.96)</td>
<td>4.05</td>
<td>&lt;.001</td>
<td>0.56</td>
<td>22 (42)</td>
</tr>
<tr>
<td><strong>Response time variability (SD of response time)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal (n=46)</td>
<td>48 (26.3)</td>
<td>85 (40.3)</td>
<td>10.3 (10.02)</td>
<td>6.95</td>
<td>&lt;.001</td>
<td>1.02</td>
<td>18 (39)</td>
</tr>
<tr>
<td>Healthy (n=53)</td>
<td>61 (31.0)</td>
<td>100 (40.0)</td>
<td>10.7 (8.64)</td>
<td>8.99</td>
<td>&lt;.001</td>
<td>1.24</td>
<td>25 (47)</td>
</tr>
<tr>
<td><strong>Commission errors (impulsivity)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal (n=46)</td>
<td>48 (26.3)</td>
<td>85 (40.3)</td>
<td>4.0 (7.23)</td>
<td>3.75</td>
<td>&lt;.001</td>
<td>0.55</td>
<td>19 (41)</td>
</tr>
<tr>
<td>Healthy (n=53)</td>
<td>61 (31.0)</td>
<td>100 (40.0)</td>
<td>2.0 (3.17)</td>
<td>4.51</td>
<td>&lt;.001</td>
<td>0.62</td>
<td>16 (30)</td>
</tr>
<tr>
<td><strong>Omission errors (inattention)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal (n=46)</td>
<td>48 (26.3)</td>
<td>85 (40.3)</td>
<td>1.5 (3.16)</td>
<td>3.27</td>
<td>.003</td>
<td>0.48</td>
<td>24 (52)</td>
</tr>
<tr>
<td>Healthy (n=53)</td>
<td>61 (31.0)</td>
<td>100 (40.0)</td>
<td>0.64 (1.88)</td>
<td>2.48</td>
<td>.02</td>
<td>0.34</td>
<td>21 (40)</td>
</tr>
</tbody>
</table>

aCPT: continuous performance task.  
bASRS: Adult ADHD Self-Report Scale.  
cADHD: attention-deficit/hyperactivity disorder.  
dADHD-RS-IV: ADHD Rating Scale IV.  
eReported in milliseconds for response time average and variability, and in number of errors for commission and omission errors.  
fReported as Benjamini-Hochberg–adjusted P values.  
gRCI: reliable change index.

EEG Indicators of Adult ADHD

**Resting EEG Baseline**

The DAR, TAR, and TBR were calculated from baseline resting EEG data (n=271) in frontal regions (ie, average of F3 and F4) with eyes closed from participants scoring in abnormal (n=125) or healthy ranges (n=146) on the adult ADHD questionnaire. Regarding the DAR, an independent-samples t test demonstrated that participants in the abnormal results group (mean 1.10, SD 0.61) had significantly greater frontal DAR than healthy participants (mean 0.90, SD 0.48; t235=3.02, P BH=.009, d=0.37). The Levene test indicated unequal variances (F=5.25, P=.02), so degrees of freedom were adjusted from 269 to 235. Post hoc independent-samples t tests confirmed that results were driven by less frontal alpha, as opposed to differences in theta (t269=1.11, P BH=.27, d=0.13).

Regarding the TAR, a comparable t test reported a significant difference for the frontal TAR (t269=2.46, P BH=.02, d=0.30), as participants with abnormal scores (mean 0.64, SD 0.30) had significantly greater ratios than those with healthy scores (mean 0.56, SD 0.26). Post hoc t tests confirmed that results were driven by less frontal alpha, as opposed to differences in theta (t269=0.53, P BH=.60, d=0.06).

Regarding the TBR, a final t test reported no significant difference between participants with abnormal scores (mean 0.66, SD 0.27) and those with healthy scores (mean 0.64, SD 0.31; t269=0.53, P BH=.60, d=0.06).

**Preliminary Resting EEG Improvement**

Changes in the DAR, TAR, and TBR in the frontal regions with eyes closed were reported for participants (n=41) scoring in the abnormal (n=20) or healthy ranges (n=21) of the adult ADHD questionnaire (Table 4). Most participants completed reduced theta (32/41, 78%), enhanced high beta (37/41, 90%), and enhanced alpha protocols (27/41, 66%). After correcting for multiple comparisons, significant improvement was only reported for the DAR in the abnormal results group.
Table 4. Change in resting EEG ratios from frontal (ie, average F3 and F4) electrodes during the eyes-closed condition after ≥30 days of Myndlift neurofeedback (n=41) for healthy users and separately for those that scored in the abnormal adult ADHD range (per conventional clinical cutoffs) at baseline.

<table>
<thead>
<tr>
<th>EEG&lt;sup&gt;a&lt;/sup&gt; pre-post outcome and group at baseline (cutoff value)</th>
<th>No. of sessions, mean (SD)</th>
<th>Treatment duration (days), mean (SD)</th>
<th>Change reduction (Hz), mean (SD)</th>
<th>Change T value</th>
<th>Change P value&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Effect size, d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta/alpha ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal (≥14; n=20)</td>
<td>49 (22.1)</td>
<td>77 (26.8)</td>
<td>0.20 (0.284)</td>
<td>3.15</td>
<td>.03</td>
<td>0.70</td>
</tr>
<tr>
<td>Healthy (&lt;14; n=21)</td>
<td>61 (36.9)</td>
<td>76 (27.8)</td>
<td>0.08 (0.450)</td>
<td>0.79</td>
<td>.59</td>
<td>0.18</td>
</tr>
<tr>
<td>Theta/alpha ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal (≥14; n=20)</td>
<td>49 (22.1)</td>
<td>77 (26.8)</td>
<td>0.04 (0.171)</td>
<td>1.00</td>
<td>.66</td>
<td>0.22</td>
</tr>
<tr>
<td>Healthy (&lt;14; n=21)</td>
<td>61 (36.9)</td>
<td>76 (27.8)</td>
<td>0.01 (0.227)</td>
<td>0.35</td>
<td>.66</td>
<td>0.08</td>
</tr>
<tr>
<td>Theta/beta ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal (≥14; n=20)</td>
<td>49 (22.1)</td>
<td>77 (26.8)</td>
<td>0.04 (0.144)</td>
<td>1.34</td>
<td>.79</td>
<td>0.30</td>
</tr>
<tr>
<td>Healthy (&lt;14; n=21)</td>
<td>61 (36.9)</td>
<td>76 (27.8)</td>
<td>0.02 (0.218)</td>
<td>0.44</td>
<td>.73</td>
<td>0.10</td>
</tr>
</tbody>
</table>

<sup>a</sup>EEG: electroencephalography.
<sup>b</sup>Reported as Benjamini-Hochberg–adjusted P values.

**Discussion**

**Principal Findings**

This retrospective study offers initial evidence of therapist-guided remote neurofeedback as an effective tool for reducing subjective symptoms, improving objective cognitive performance, and adaptively modifying EEG markers. Improvements in attention were evident in children and adults with ADHD, as well as healthy participants. Findings suggest that the TBR is not a reliable marker for adult ADHD, instead demonstrating alternative elevated slow/fast power ratios [37]. Moreover, we provide preliminary evidence for improvement (ie, reduced DAR) in adults with ADHD. These findings offer a promising use for remote NFT as a low-cost alternative to clinic-based EEG.

**Efficacy for Improving Mental Health Remotely**

Based on real-world data, significant improvement was reported across standardized questionnaires. The greatest improvement was observed in participants with abnormal anxiety scores, where most received reduced theta, enhanced high beta, and enhanced alpha protocols. As anticipated, greater effect sizes were observed for participants with scores in the abnormal versus healthy ranges. Interestingly, healthy participants and those with ADHD, both children and adults, demonstrated significant improvement with large effect sizes after completing primarily reduced theta and enhanced high beta protocols, as well as adults who completed enhanced alpha protocols or children who completed reduced low beta and reduced SMR protocols. Consistent with the literature [33-35], our findings suggest that children and adults may benefit from unique NFT protocols to improve ADHD symptoms, although a larger sample is required to confirm preliminary ADHD-RS-IV results.

Apart from the child ADHD assessment, questionnaire analyses included large total sample numbers (ie, 112 to 263 participants), and after an average of 53 NFT sessions, 57% to 78% of the participants demonstrated significant improvement, depending on the questionnaire. Results were particularly impressive compared to other in-app mental health therapeutics [73-76], such as mobile-enabled text psychotherapy [77] or app-based cognitive behavioral therapy [78]. The majority (61%) of participants scoring in the abnormal ranges moved to the healthy results group over an average of approximately 3 months, a timeframe costing less than US $1500 with Myndlift versus US $6000 to $8000 for traditional neurofeedback.

**Improved Cognitive Performance for Healthy Participants and Those With ADHD**

NFT led to greater consistency in response times on a response inhibition task for subjects scoring in healthy or abnormal ADHD ranges, agreeing with similar reports of subjects with ADHD [79,80]. In addition, the RCI demonstrated that approximately 50% of healthy participants improved their average response time, while similarly, participants in abnormal ranges reduced omission errors. Importantly, CPT findings agree with improved ADHD questionnaire scores, suggesting that NFT provides objective evidence of improved executive function, the primary cognitive domain impacted by attentional difficulties.

**Identifying Adult ADHD Neuromarkers**

Resting EEG findings demonstrated that elevated DAR and TAR were indicative of adult ADHD at baseline. This translated to significantly higher levels of delta and lower levels of alpha, as previously reported in adults with ADHD [39-42,81,82]. Notably, Liechti and colleagues [35] reported high theta to be less consistent in adults than in children, and that ADHD versus healthy control classification improved having exploratorily included delta waves in the discriminant analysis. Adults with ADHD may present slower theta waves—bordering fast delta waves—than children, although further analysis is required. Together, findings are consistent with the cortical hypoarousal theory, where low-power fast oscillations accompany reduced self-control and executive functioning [83], and high-power slow oscillations are reported with decreased subcortical motivational drive [84]. Preliminary evidence for reduced DAR
in adult ADHD from pre- to postassessment may reflect the improved ADHD symptoms and CPT measures, particularly given the success of protocols inhibiting slow oscillations and enhancing fast oscillation [1], and the high percentage of ADHD participants performing reduced theta (ie, slow) and enhanced alpha (ie, fast) protocols.

In contrast to our work and that of others, several groups reported high alpha power at baseline during eyes-closed conditions in adult ADHD populations [85,86], or rather, no difference across ADHD participants and healthy controls [87,88]. Importantly, variability across the adult ADHD literature may, in part, be due to the heterogeneity of ADHD [34,35] and differences in study designs, sample sizes, analyses, and EEG technology [89]. For example, Loo and colleagues [38] demonstrated that adults with ADHD combined-type (ie, symptoms of inattention and hyperactivity or impulsivity) present reduced alpha power globally, compared to ADHD inattentive-type or non-ADHD controls.

Limitations and Future Directions
Study results are encouraging, but conclusions should be tempered by limitations, including small subgroup sample sizes and lack of control groups. Moreover, subjects may have received alternative treatment in parallel (eg, medication) that could influence symptom improvement as well as alter neuromarkers. For example, two studies administering stimulants (ie, methylphenidate or dexamphetamine) to treat symptoms of ADHD in adults demonstrated altered delta [90,91] and theta waves [90] posttreatment. No changes in alpha or beta waves were reported. Given the evidence in this study for altered delta and alpha waves in adults with abnormal ADHD scores, we would hypothesize that the mechanism of action for stimulants versus NFT may differ, resulting in influence over varied frequency bands. Moreover, as this population reflects real-world use, the likelihood of these two forms of treatment to have commenced simultaneously, for treating symptoms of depression, anxiety, and ADHD, would arguably be low. Those seeking treatment with remote neurofeedback most often do so to avoid taking pharmaceuticals [92,93] or, rather, to supplement their current treatment, which alone may not be sufficiently effective [94]. Frank H Duffy [95], a Harvard professor and pediatric neurologist, suggests that “if any medication had demonstrated such a wide spectrum of efficacy it would be universally accepted and widely used.” Further, controlled research studies will be required to facilitate comparison of neurofeedback efficacy with other interventions. Notwithstanding these limitations, the findings are essential as they reflect real-world benefits of remote neurofeedback to actual patients. Follow-up analyses will compare benefits across NFT protocols and will further evaluate the impact on resting EEG outcomes.

Conclusions
Preliminary findings from this retrospective pilot study demonstrate efficacy of remote NFT in improving mental health, particularly for individuals with symptoms of ADHD and anxiety, mainly through reduced theta, enhanced high beta, and enhanced alpha NFT protocols. Moreover, adult ADHD was distinguished from healthy individuals by elevated frontal DARs, where ratios were significantly reduced following NFT. The effectiveness of the system in a real-world population via remote use positions it as an affordable and accessible alternative to clinic-based systems.

Acknowledgments
Funding for this study was provided by McGill University, Montreal, Canada, and Myndlift Ltd. JCW received funding from McGill University for a doctoral internship.

Authors’ Contributions
JCW conducted the analyses, interpreted the results, cowrote the paper, and revised the final version. RN co-designed the study, extracted the data, interpreted the results, cowrote the paper, and revised the final version. GMD designed the study, cowrote the paper, and revised the final draft. All authors reviewed and revised the results and approved the final version of the paper.

Conflicts of Interest
JCW is a freelance consultant for Myndlift. RN and GMD are employees of Myndlift.

Multimedia Appendix 1
Neurofeedback protocols implemented in this study.
[DOCX File , 18 KB - formative_v6i7e35636_app1.docx ]

Multimedia Appendix 2
Electroencephalography preprocessing.
[DOCX File , 13 KB - formative_v6i7e35636_app2.docx ]

Multimedia Appendix 3
Continuous performance task improvement for users without an ADHD assessment. ADHD: attention-deficit/hyperactivity disorder.
References


Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ADHD-RS-IV: ADHD Rating Scale IV
AF: anterior frontal
ASRS: Adult ADHD Self-Report Scale
BH: Benjamini-Hochberg
C: central
CPT: continuous performance task
DAR: delta/alpha ratio
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
EEG: electroencephalography
F: frontal
GAD-7: 7-item Generalized Anxiety Disorder scale
GHQ-12: 12-item General Health Questionnaire
IRB: Institutional Review Board
NFT: neurofeedback training
O: occipital
P_{BH}: Benjamini-Hochberg–adjusted $P$ value
PHQ-9: 9-item Patient Health Questionnaire
RCI: reliable change index
RT: response time
SDRT: SD of response time, response time variability
SMR: sensorimotor rhythm
TAR: theta/alpha ratio
TBR: theta/beta ratio
TP: temporal pole
z: zero
Novel Implementation Strategy to Electronically Screen and Signpost Patients to Health Behavior Apps: Mixed Methods Implementation Study (OptiMine Study)

Zarnie Khadjesari¹, PhD; Tracey J Brown¹, MSci; Alex T Ramsey², PhD; Henry Goodfellow³, PhD; Sherine El-Toukhy⁴, PhD; Lorien C Abroms⁵, PhD; Helena Jopling⁶, MBBS, MPhil; Arden Dierker Viik⁶, MS; Michael S Amato⁷, PhD

¹Behavioural and Implementation Science Research Group, School of Health Sciences, University of East Anglia, Norwich, United Kingdom
²Department of Psychiatry, Washington University School of Medicine, St Louis, MO, United States
³Department of Primary Care and Population Health, University College London, London, United Kingdom
⁴Division of Intramural Research, The National Institute on Minority Health and Health Disparities, The National Institutes of Health, Bethesda, MD, United States
⁵Department of Prevention and Community Health, Milken Institute School of Public Health, The George Washington University, Washington DC, DC, United States
⁶Department of Public Health, West Suffolk NHS Foundation Trust, Bury St Edmunds, United Kingdom
⁷Truth Initiative, Washington DC, DC, United States

Corresponding Author:
Zarnie Khadjesari, PhD
Behavioural and Implementation Science Research Group
School of Health Sciences
University of East Anglia
Norwich Research Park
Norwich, NR4 7TJ
United Kingdom
Phone: 44 01603457648
Email: Z.Khadjesari@uea.ac.uk

Abstract

Background: Behavior change apps have the potential to provide individual support on a population scale at low cost, but they face numerous barriers to implementation. Electronic health records (EHRs) in acute care hospitals provide a valuable resource for identifying patients at risk, who may benefit from behavior change apps. A novel, emerging implementation strategy is to use digital technologies not only for providing support to help-seeking individuals but also for signposting patients at risk to support services (also called proactive referral in the United States).

Objective: The OptiMine study aimed to increase the reach of behavior change apps by implementing electronic signposting for smoking cessation and alcohol reduction in a large, at-risk population that was identified through an acute care hospital EHR.

Methods: This 3-phase, mixed methods implementation study assessed the acceptability, feasibility, and reach of electronic signposting to behavior change apps by using a hospital’s EHR system to identify patients who are at risk. Phase 1 explored the acceptability of the implementation strategy among the patients and staff through focus groups. Phase 2 investigated the feasibility of using the hospital EHR to identify patients with target risk behaviors and contact them via SMS text message, email, or patient portal. Phase 3 assessed the impact of SMS text messages sent to patients who were identified as smokers or risky drinkers, which signposted them to behavior change apps. The primary outcome was the proportion of participants who clicked on the embedded link in the SMS text message to access information about the apps. The acceptability of the SMS text messages among the patients who had received them was also explored in a web-based survey.

Results: Our electronic signposting strategy—using SMS text messages to promote health behavior change apps to patients at risk—was found to be acceptable and feasible and had good reach. The hospital sent 1526 SMS text messages, signposting patients to either the National Health Service Smokefree or Drink Free Days apps. A total of 13.56% (207/1526) of the patients clicked on the embedded link to the apps, which exceeded our 5% a priori success criterion. Patients and staff contributed to the SMS text message content and delivery approach, which were perceived as acceptable before and after the delivery of the SMS text.
messages. The feasibility of the SMS text message format was determined and the target population was identified by mining the EHR.

**Conclusions:** The OptiMine study demonstrated the proof of concept for this novel implementation strategy, which used SMS text messages to signpost at-risk individuals to behavior change apps at scale. The level of reach exceeded our a priori success criterion in a non-help-seeking population of patients receiving unsolicited SMS text messages, disconnected from hospital visits.

**International Registered Report Identifier (IRRID):** RR2-10.2196/23669

**KEYWORDS**
electronic health record; EHR; alcohol reduction; electronic messages; proactive messages; proactive outreach; smoking cessation; tobacco use; alcohol use; alcohol; smoking; mobile health; mHealth; mobile app

**Introduction**

More than one-third of cancers are preventable by adopting a healthier lifestyle, such as stopping smoking and reducing risky drinking; tackling these health behaviors would lead to significant health benefits, minimize comorbidity, and reduce the burden on the National Health Service (NHS) [1-3]. The traditional approach to health promotion in the hospital environment is centered on health professionals providing verbal advice or information in leaflets, with or without the offer of referral to a lifestyle service. This approach relies on the clinician’s knowledge, skills, and confidence; time; and the availability of health promotion literature, of which all are significant barriers to delivery [4,5]. Attempts have been made to boost health promotion activities in acute care hospitals in England. For example, (1) financial incentives have been provided for hospitals that undertake screening and brief advice for smoking and risky drinking (2018-2020) [6]; (2) brief advice on health behaviors through opportunistic day-to-day interactions between patients and health care professionals is promoted via Making Every Contact Count, an NHS initiative [7]; and (3) public health specialists are increasingly being embedded into the acute care context in recognition of the links between health behaviors and chronic illnesses, such as cancer, and the pressure this puts on the NHS. However, these interventions are typically resource-intensive or have been discontinued owing to resource constraints.

Digital interventions, such as behavior change websites or apps, have been developed to bridge the evidence-to-practice gap. They can be used as an alternative or adjunct to face-to-face delivery and have a substantial evidence base as interventions [8-12]. Digital interventions can provide effective individual support on a population scale at low cost [13]. They overcome barriers to delivering face-to-face interventions, such as removing the stigma associated with seeking help and alleviating pressure on busy health professionals to deliver brief interventions. Digital interventions for addictive behaviors are more commonly offered in higher education, primary care, and community settings (although far from routine) but rarely in the context of acute health services [8,9,14-16]. Furthermore, the implementation of effective digital interventions is often not considered and relies on the help-seeking behavior of motivated individuals. Another underused digital resource in the acute care setting is the electronic health record (EHR) for screening patients at risk. Although screening is commonplace for promoting medication adherence, vaccine uptake, or cancer screening, systematic screening for health behaviors (such as risky alcohol consumption and tobacco smoking) for signposting to support services is in its infancy.

We developed a novel implementation strategy that uses the combined potential of three existing technologies to bridge the evidence-to-practice gap: (1) acute care hospital EHR for identifying at-risk individuals at scale, (2) health promotion apps and websites that provide behavior change support at an individual level, and (3) SMS text messages that are commonly used by health services to communicate with patients. Using electronic messages (such as SMS text message or email) to signpost (referred to as *proactive outreach* in the United States) patients who are at risk, identified by the EHR, to health promotion apps is cheap, efficient, quick to implement using existing infrastructure, and scalable to other health behaviors and health care settings and has the potential to achieve behavior change at the population level. Although SMS text messaging has been used successfully as a treatment tool for smoking cessation [11], using it as a primary channel for outreach and enrollment to support services or treatment represents a novel application with preliminary success. Krebs et al [18] used outreach SMS text messages in a large New York health system to connect patients to quitline counseling. Furthermore, Abroms et al [20] used SMS text messaging to connect patients in the emergency department to a smoking cessation SMS text messaging program or quitline counseling.

So far, no previous studies have evaluated the ability of electronic messages to signpost patients who are not seeking help to behavior change apps, disconnected from hospital visits. The aim of this 3-phase, mixed methods implementation study (the OptiMine study) was to explore the acceptability (phase 1), feasibility (phase 2), and reach (phase 3) of electronic messages to signpost patients who smoke tobacco and drink alcohol at risky levels to behavior change apps. The primary outcome of phase 1 was a qualitative synthesis of patient-perceived attributes of signposting. The primary outcome of phase 2 was the estimated size and characteristics of the population that may be reached by the intervention. Findings from these 2 phases informed the design of phase 3, the primary outcome of which was the proportion of contacted patients who followed the signpost to access more information about the promoted behavior change apps (ie, the *click rate*). The approach and methods have been reported in our published protocol [21].
Methods

Ethics Approval
The NHS Research Ethics Committee and the Health Research Authority approved this study in June 2019. EHR data were accessed and analyzed by staff within the West Suffolk NHS Foundation Trust (WSFT) information service team. Patients at risk were identified by the WSFT information service team and received electronic messages directly from the hospital. All analyses were conducted by the WSFT information service and public health teams. The study team members had access only to anonymized, aggregated data. Authorization for using EHR data at the WSFT was provided by the Information Governance Team via Data Protection Impact Assessment (approval date: September 20, 2019) [22].

Theoretical Frameworks
We used the taxonomy of implementation outcomes by Proctor et al [23] to design this implementation study, focusing on acceptability, feasibility, and reach. Acceptability and feasibility are important precursors for effective uptake and reach of an intervention and were explored using qualitative focus groups (acceptability before implementation), a web-based survey (acceptability after implementation), and routinely collected data from the EHR (feasibility). We assessed reach as our measure of implementation success, operationalized as the proportion of patients who engaged with our SMS text message by clicking on an embedded link to access free apps to support health behavior change. The topic guides and survey exploring the acceptability of delivering electronic messages to patients were informed by the Perceived Attributes of eHealth Innovations [24]—an extension of the Diffusion of Innovations Theory [25], which has been applied through a validated questionnaire to test the acceptability of a digital innovation [26]. Further information on the model and its application in this study is provided in our published protocol [21].

Setting
The research was conducted between April 2019 and July 2020 within the West Suffolk Hospital, an acute NHS provider renowned for its world-leading delivery of care using digital technologies, that is, an acute Global Digital Exemplar Trust [27]. eCare (trade name: Cerner Millenium), the EHR system at the WSFT, was launched in 2016 and includes records for all outpatients and inpatients registered with the hospital. The EHR contains contact information, demographics, and health data such as chronic disease status needed for this study. Data on smoking and alcohol consumption status were collected as part of the lifestyle screening survey or the Activities of Daily Living assessment, routinely provided to patients on admission to the hospital. The hospital is situated in the rural region of West Suffolk in the East of England, where population characteristics are similar to those of the general population in England, with slightly higher proportions of individuals aged >65 years and White British residents [28].

Implementation Strategy and Behavior Change Apps
We electronically signposted patients to the NHS Smokefree and Drink Free Days apps. Public Health England has developed the Smokefree and the Drink Free Days apps as part of their One You campaign for supporting healthy lifestyles. These apps are freely available on the web [29] and are heavily promoted in the United Kingdom mass media. The apps are theoretically informed and use evidence-based behavior change techniques, such as goal setting and self-monitoring. This mixed methods implementation study explored acceptability (phase 1) and feasibility (phase 2) as important precursors for the reach of our implementation strategy (phase 3).

Phase 1: Acceptability of Electronic Signposting (Before Implementation)
Acceptability of the implementation strategy was explored before its delivery via focus groups. Patients and staff were invited to participate in face-to-face focus groups to explore their perspectives. Patients were eligible if they were smoking or drinking alcohol regularly. Eligible staff were those (1) in senior IT management roles, (2) responsible for administering lifestyle screening, or (3) involved in EHR data management and hospital communications. Patients were identified via volunteer coordinators, a news story on the hospital website, and a recruitment stall at the main hospital entrance. Staff were recruited directly through email invitation and a weekly staff newsletter. All participants were provided with a participant information sheet and a consent form. Patient focus groups were conducted on-site at the education center at WSFT, whereas staff focus groups were conducted in meeting rooms. Focus groups were audio-recorded and transcribed verbatim by a professional transcription company, removing any identifiable data. Transcripts were coded using NVivo (QSR International). Framework analysis was used to synthesize the findings of the focus groups, based on 3 domains of the Perceived Attributes Theory: compatibility (ie, degree to which electronic messaging was consistent with patient preferences), complexity (ie, degree to which electronic messaging was difficult to understand or act on), and relative advantage (ie, degree to which electronic messaging was superior to alternative or more traditional approaches) [24]. Subthemes were focused on the pragmatic development, refinement, and delivery of the implementation strategy.

Phase 2: Feasibility of Electronic Signposting
Feasibility of using the EHR to identify patients at risk who have mobile phone numbers, email addresses, and patient portal access was explored through data mining. Patients with alcohol consumption or smoking status that had been recorded or updated within the past 13 months were included. The Activities of Daily Living and customized lifestyle screening assessments are used by the WSFT to record alcohol consumption and smoking status on admission to the hospital. The following data were extracted and aggregated from the EHR by a hospital-based information analyst: smoking status (yes or no), alcohol consumption status: Alcohol Use Disorders Identification Test–Consumption (AUDIT-C) score [30] (low risk: 0–4, at risk: 5–9, and dependent: 10–12), sex (woman or man), mobile phone number (yes or no), email address (yes or no), and patient portal access (yes or no). Frequencies and percentages of patients with valid data in each of the fields mentioned...
previously were reported to the study team, who had no direct access to the underlying data.

**Phase 3: Reach of Electronic Signposting**

**Participants**

Eligible patients were adults (aged ≥18 years) recorded as smoking or drinking alcohol at risky levels within the past 13 months, with a valid mobile number. Patients were excluded if they were pregnant, were registered on the end-of-life pathway, or had opted out of communications from the hospital. As specified in the protocol, a minimum sample size of 383 per risk profile group would allow calculation of 95% CIs within a margin of −5% to +5% points, assuming a population proportion of 50% and a population size of 100,000 [21].

**Procedure**

SMS text messages were selected as the most acceptable and feasible electronic format based on the findings of phase 1 and phase 2. Although the protocol sought to compare 3 risk profiles (exclusive smokers, exclusive risky drinkers, and both), to reach the a priori specified minimum sample size, it was necessary to collapse to two risk profiles: (1) exclusive smokers and (2) risky drinkers regardless of smoking status. The hospital’s SMS text messaging system was used to send an initial message to all the participants, signposting to either the Smokefree or Drink Free Days apps based on risk profile, with a second reminder message sent 3 days later to any participant who had not clicked the link yet. Unique link URLs were used to identify the participants who clicked the link. Figure 1 illustrates the content and delivery schedule of the messages.

**Figure 1.** Flowchart of SMS text message content and delivery.
Data Collection

To determine the characteristics of patients to whom SMS text messages were sent, data were retrieved from eCare, and analyses were performed by a public health manager with the support of coauthors MSA and ZK in July 2020. The following variables were extracted from the system and categorized as follows:

1. Name (for data linkage purposes only)
2. Hospital and NHS number (for data linkage purposes only)
3. Date of birth (for data linkage purposes only)
4. Age—categorized as 18-25, 26-35, 36-45, 46-55, 56-65, and >66 years
5. Sex—woman or man
6. Ethnicity—categories were merged to increase patient numbers in less-populated groups: any White background, any other background, and not known or not stated; refer to the study protocol for individual categories [21]
7. Postcode—used to calculate index of multiple deprivation decile—grouped into quintiles
8. Date of smoking or alcohol consumption screening at the hospital—categorized as days since screening: 1-4, 4-7, 7-10, 10-13, and >13 months
9. SMS text message sent, targeting smoking cessation or alcohol reduction
10. Embedded link to apps clicked or not clicked

Indices of Multiple Deprivation

Postcodes were entered into the government multiple deprivation lookup to obtain the deciles of multiple deprivation for each patient [31], where decile 1 represents the most deprived 10% of the population and decile 10 represents the least deprived 10% of the population.

Recency of Screening

The number of days since patients were screened for smoking or alcohol consumption was calculated by subtracting the date of smoking or alcohol screening from the date on which the smoking or alcohol SMS text message was sent. This was rounded to a whole number.

Health Data

The long-term health conditions stored as structured data in eCare were audited against the whole medical record to determine its suitability for use. Auditors used inpatient, emergency department, and general practice notes to collate patients’ medical history. More than half of the manually audited records found different health data than those retrieved from the patient records. Thus, the health data were considered too incomplete and inaccurate to be included in the study. Refer to the study protocol for the list of long-term health conditions originally intended for extraction from eCare [21].

Missing Data in AUDIT-C Fields

A large proportion of the records (1424/1975, 72.1%) that documented alcohol status had missing data in the AUDIT-C fields. Scores were calculated from the available fields, imputing a value of zero for missing fields, based on information from hospital staff that fields were most likely skipped because they were not applicable (eg, erroneously left blank instead of selecting 0). Patients with missing data and calculated AUDIT-C scores of 7 to 8 were excluded from the study because it was not possible to be confident that these patients were not dependent drinkers, for whom the intervention would be clinically inappropriate.

Statistical Modeling

We determined a priori that 5% reach, as evidenced by clicking on the embedded links to the apps, would constitute a clinically meaningful level of reach, given the low-burden and scalable nature of the interventions [21]. This success criterion was based on the rationale that reaching even this modest proportion of non-help-seeking individuals with high-risk drinking and smoking behaviors via unsolicited SMS text messages at a time disassociated with their last hospital visit could have great impact at a population health level. The number of patients who clicked the embedded link within their respective SMS text message was reported along with their baseline characteristics. The denominator for reach rate was the total number of patients with that characteristic who received an SMS text message. R software (R Foundation for Statistical Computing) was used to model the logistic regression, where we used the glm function with a Poisson distribution. To investigate potential relationships of common demographic covariates and assess any effect of screening recency, the model included all patient-level variables that were available: sex, age group, ethnicity group, index of multiple deprivation quintile, and days since screening group. Relative risks (RRs) with 95% CIs for each characteristic were derived from the model in R, using exponentiated values of the model coefficients.

Acceptability of Electronic Signposting (After Implementation)

Acceptability of the implementation strategy was also explored after its delivery via a web-based survey. Eligible patients were those to whom a signpost SMS text message was sent. Then, 6 days after the initial signpost SMS text message was sent, 2 additional SMS text messages were sent to the participants (an initial message and a reminder), inviting them to participate in a web-based survey about their views on receiving the signpost SMS text messages. Participant information sheets and consent forms were incorporated into the web-based survey. We used JISC web-based surveys, a free web-based platform designed for academic research and public sector organizations. Participants were given 15 days to complete the survey.

Results

Phase 1: Acceptability of Electronic Signposting (Before Implementation)

A total of 10 patients participated in 2 focus groups (group 1: n=3, 30% of the participants; group 2: n=6, 60% of the participants) and 1 individual interview (n=1, 10% of the participants; owing to low turnout for the focus group). A total of 14 staff members participated in 3 focus groups (group 1: n=5, 36% senior managers; group 2: n=1, 7% nurse and n=1, 7% pharmacy technician; group 3: n=7, 50% members of IT staff and communications officers). Patients’ ages were collected as categories and ranged from 46 to >66 years; 60% (6/10) were
women. Most patients (7/10, 70%) were members of the hospital’s volunteer group and patient portal user group, with 20% (2/10) of them being members of the public and 10% (1/10) being members of staff identified via the recruitment stall. Findings from this phase suggested that most patients found SMS text messaging as the most acceptable form of electronic message for receiving communications. A more detailed summary of the key findings under each theme and subtheme is presented in Multimedia Appendix 1 (qualitative focus group findings), along with illustrative quotes and the approach that informed the development and delivery of electronic messages.

Phase 2: Feasibility of Electronic Signposting

The dynamic nature of an acute care hospital EHR database was found to be an implementation challenge. The mining of data from eCare needed to be performed on different days, owing to the size of the data queries, and with different search strategies depending on the fields required. Furthermore, the results changed depending on the date on which the queries were run (both the denominators and numerators). Queries that were intended to generate data sets that would include the whole EHR population could only be achieved by limiting the number of variables in the data set; large queries often took several hours to run and were at risk of timing out before they had completed.

The total number of adult patients in the eCare system on October 9, 2019, was 228,982, of which 1092 (0.48%) adults were recorded as smokers (smoking and drinking status data extracted on January 10, 2020). Among the 0.74% (1702/228,982) of the patients who were recorded as drinking alcohol, 23.62% (402/1702) patients were recorded as at-risk drinkers. An additional 0.55% (1249/228,982) and 0.02% (51/228,982) of the patients were reported as drinking at low risk or dependent levels, respectively. Most patients (226,784/228,982, 99.04%) had missing data in the EHR regarding their tobacco or alcohol use status. Smoking and drinking status are only recorded in a way that is retrievable when a patient has an inpatient admission. The proportion of the catchment population that is admitted each year is typically 13% [32]. The proportions of admitted patients who were screened for smoking and alcohol use in the financial year 2018 to 2019 were reported by WSFT information service staff to be 64% and 68%, respectively.

Of the 228,982 patients, 146,171 (63.84%) had mobile phone numbers, of which 707 (0.48%) were recorded as smokers and 271 (0.19%) were recorded as at-risk drinkers. Of the 228,982 patients, the total number of patients with email addresses was 32,375 (14.14%), of which 137 (0.42%) were recorded as smokers and 115 (0.36%) were recorded as at-risk drinkers. The database of patient portal users was independent of the EHR, and could not be linked to the EHR by the WSFT information service staff. The numbers of patient portal users recorded as smoking and drinking could not be determined. Therefore, the consideration of the patient portal as a form of message delivery was discontinued in this study. Combined with the findings from phase 1, these feasibility findings from phase 2 reinforced the decision to use SMS text messages as the channel for the electronic messages.

Phase 3: Reach of Electronic Signposting

Baseline Characteristics and Risk Profile Groups

On the basis of the findings in phase 2, the participant sample was mined from eCare on January 10, 2020. The sampling frame used the most recent admissions to the hospital with validated data to identify patients whose records were most likely to be up to date and accurate. A 13-month time frame (October 1, 2018, to November 30, 2019) was used to obtain a sufficiently large population of patients to meet the minimum sample size required, which was balanced against recency of the data. A total of 6521 people admitted during the time frame were aged ≥18 years, not on the end-of-life pathway, and not pregnant and had a mobile phone number recorded and either their smoking or alcohol status recorded. Owing to the relatively small number of patients who could be identified as exclusively consuming alcohol at risky levels but who were nonsmokers, all participants with risky drinking were combined into a single risk profile group, regardless of their smoking status. The resulting 2 risk profile groups (smoking only and risky drinking with or without smoking) represent a deviation in analysis from the protocol, which aimed to create 3 risk profile groups (smoking only, risky drinking only, and smoking and risky drinking). This deviation was necessary to meet our a priori minimum sample size of 383 participants per group, which was also specified in the protocol. The decision to collapse the risky drinking groups was based exclusively on the baseline data, before the analyses of outcomes. Ultimately, of the 1526 individuals, the selected sample included 1103 (72.28%) individuals who were recorded as smokers only, 276 (18.09%) individuals who were recorded as risky drinkers only, and 147 (9.63%) individuals who were recorded as both smokers and risky drinkers. Table 1 shows the baseline characteristics of the participants.
Table 1. Baseline characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Smoker only&lt;sup&gt;a&lt;/sup&gt; (n=1103)</th>
<th>Risky drinker with or without also being a smoker&lt;sup&gt;b&lt;/sup&gt; (n=423)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>553 (50.14)</td>
<td>129 (30.5)</td>
</tr>
<tr>
<td>Men</td>
<td>550 (49.86)</td>
<td>294 (69.5)</td>
</tr>
<tr>
<td>Not recorded</td>
<td>N/A&lt;sup&gt;c&lt;/sup&gt;</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-25, n (%)</td>
<td>124 (11.24)</td>
<td>17 (4)</td>
</tr>
<tr>
<td>26-35, n (%)</td>
<td>200 (18.13)</td>
<td>29 (6.9)</td>
</tr>
<tr>
<td>36-45, n (%)</td>
<td>192 (17.41)</td>
<td>60 (14.2)</td>
</tr>
<tr>
<td>46-55, n (%)</td>
<td>205 (18.59)</td>
<td>90 (21.3)</td>
</tr>
<tr>
<td>56-65, n (%)</td>
<td>176 (15.96)</td>
<td>98 (23.2)</td>
</tr>
<tr>
<td>66-75, n (%)</td>
<td>125 (11.33)</td>
<td>78 (18.4)</td>
</tr>
<tr>
<td>&gt;75, n (%)</td>
<td>81 (7.34)</td>
<td>51 (12.1)</td>
</tr>
<tr>
<td>Not recorded</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1021 (92.57)</td>
<td>395 (93.4)</td>
</tr>
<tr>
<td>Mixed</td>
<td>2 (0.18)</td>
<td>1 (0.2)</td>
</tr>
<tr>
<td>Asian or Asian British</td>
<td>2 (0.18)</td>
<td>1 (0.2)</td>
</tr>
<tr>
<td>Black or Black British</td>
<td>5 (0.45)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Other ethnic groups</td>
<td>21 (1.90)</td>
<td>8 (1.9)</td>
</tr>
<tr>
<td>Not recorded</td>
<td>52 (4.71)</td>
<td>18 (4.3)</td>
</tr>
<tr>
<td><strong>Index of multiple deprivation (quintile), n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>77 (6.98)</td>
<td>17 (4)</td>
</tr>
<tr>
<td>2</td>
<td>279 (25.29)</td>
<td>92 (21.7)</td>
</tr>
<tr>
<td>3</td>
<td>371 (33.64)</td>
<td>147 (34.8)</td>
</tr>
<tr>
<td>4</td>
<td>251 (22.76)</td>
<td>111 (26.2)</td>
</tr>
<tr>
<td>5</td>
<td>121 (10.97)</td>
<td>49 (11.6)</td>
</tr>
<tr>
<td>Not recorded</td>
<td>4 (0.36)</td>
<td>7 (1.7)</td>
</tr>
<tr>
<td><strong>Recency of screening data (months), n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-1</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>1-3</td>
<td>119 (10.79)</td>
<td>54 (12.8)</td>
</tr>
<tr>
<td>3-6</td>
<td>249 (22.57)</td>
<td>127 (30)</td>
</tr>
<tr>
<td>6-12</td>
<td>485 (43.97)</td>
<td>209 (49.4)</td>
</tr>
<tr>
<td>&gt;12</td>
<td>250 (22.67)</td>
<td>33 (7.8)</td>
</tr>
<tr>
<td>Not recorded</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

<sup>a</sup>Group includes all participants recorded as smoking and not drinking alcohol at risky levels.

<sup>b</sup>Group includes all participants recorded as drinking alcohol at risky levels, regardless of smoking status, owing to sample size considerations. The risky drinker status is defined as an Alcohol Use Disorders Identification Test–Consumption score of 5 to 10 if 3 of its items were completed or a score of 5 to 6 if only 2 items were completed.

<sup>c</sup>N/A: not applicable.

---

**Reach**

In January 2020, an SMS text message was sent to 1526 patients, signposting them to either the NHS Smokefree (n=1103, 72.28%) or the Drink Free Days apps (n=423, 27.72%). A total of 13.56% (207/1526) of the participants clicked on the embedded link to the apps (smokers: 26/207, 12.56% and risky drinkers: 34/207, 16.43%), which exceeded our 5% a priori
success criterion in both groups. **Figure 2** shows a CONSORT (Consolidated Standards of Reporting Trials) diagram. Characteristics of patients who clicked the embedded link within the SMS text message versus the characteristics of those who did not click the embedded link are presented in **Table 2**. The only significant differences in baseline characteristics were observed among smokers. Among smokers, the lowest click rate was among participants aged ≥66 years (6.8%). Compared with this age group as reference, smoking participants aged 36 to 45 years (18.2%; RR=2.6; 95% CI 1.4-4.6) and those aged 56 to 65 years (13.1%; RR=1.98; 95% CI 1.05-3.73) were significantly more likely to click. Male smokers were significantly less likely to click than female smokers (10.2% vs 14.8%; RR=0.71; 95% CI 0.51-0.98). Although no significant differences by age were observed among risky drinkers, a numerically different distribution emerged, such that risky drinkers aged 56 to 65 years were the most likely to click the link (20.4%).

**Figure 2.** CONSORT (Consolidated Standards of Reporting Trials) diagram. *Alcohol Use Disorders Identification Test–Consumption (AUDIT-C) score of 5-10 if 3 fields populated or score of 5-6 if 2 fields populated; **AUDIT-C score<5; ***both=patients who are both smokers and risky drinkers.
### Table 2. Comparison of characteristics of patients who clicked versus those who did not click on links to apps within the SMS text message.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Smokers—patients who clicked versus those who did not click&lt;sup&gt;a&lt;/sup&gt; (n=1103)</th>
<th>Risky drinkers—patients who clicked versus those who did not click&lt;sup&gt;b&lt;/sup&gt; (n=423)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patients who clicked, n (%)</td>
<td>RR&lt;sup&gt;c&lt;/sup&gt; (95% CI)</td>
</tr>
<tr>
<td>Overall click rate</td>
<td>138 (12.51)</td>
<td>—</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>82 (14.83)</td>
<td>Reference</td>
</tr>
<tr>
<td>Men</td>
<td>56 (10.18)</td>
<td>0.71 (0.51-0.98)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-25</td>
<td>14 (11.29)</td>
<td>1.55 (0.76-3.13)</td>
</tr>
<tr>
<td>26-35</td>
<td>26 (13)</td>
<td>1.86 (1.3-4.44)</td>
</tr>
<tr>
<td>36-45</td>
<td>35 (18.23)</td>
<td>2.56 (1.42-4.64)</td>
</tr>
<tr>
<td>46-55</td>
<td>26 (12.68)</td>
<td>1.83 (0.99-3.39)</td>
</tr>
<tr>
<td>56-65</td>
<td>23 (13.07)</td>
<td>1.98 (1.05-3.73)</td>
</tr>
<tr>
<td>&gt;66</td>
<td>14 (6.80)</td>
<td>2.12 (1.20-3.75)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>120 (11.75)</td>
<td>Reference</td>
</tr>
<tr>
<td>People of color</td>
<td>6 (20)</td>
<td>1.63 (0.74-3.57)</td>
</tr>
<tr>
<td>Not known or not stated</td>
<td>12 (23.08)</td>
<td>2.12 (1.20-3.75)</td>
</tr>
<tr>
<td><strong>Postcode (for index of multiple deprivation; quintile)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>10 (12.99)</td>
<td>0.96 (0.45-2.04)</td>
</tr>
<tr>
<td>2</td>
<td>37 (13.26)</td>
<td>0.96 (0.55-1.68)</td>
</tr>
<tr>
<td>3</td>
<td>49 (13.21)</td>
<td>0.96 (0.56-1.65)</td>
</tr>
<tr>
<td>4</td>
<td>26 (10.36)</td>
<td>0.76 (0.42-1.37)</td>
</tr>
<tr>
<td>5</td>
<td>16 (13.22)</td>
<td>Reference</td>
</tr>
<tr>
<td><strong>Recency of screening data (months)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-4</td>
<td>28 (13.93)</td>
<td>1.11 (0.66-1.88)</td>
</tr>
<tr>
<td>4-7</td>
<td>29 (11.20)</td>
<td>0.88 (0.52-1.47)</td>
</tr>
<tr>
<td>7-10</td>
<td>34 (14.72)</td>
<td>1.18 (0.71-1.96)</td>
</tr>
<tr>
<td>10-13</td>
<td>24 (10.08)</td>
<td>0.80 (0.46-1.38)</td>
</tr>
<tr>
<td>&gt;13</td>
<td>23 (13.22)</td>
<td>Reference</td>
</tr>
</tbody>
</table>

<sup>a</sup>These messages were sent to patients recorded as smoking and not recorded as drinking alcohol at risky levels.

<sup>b</sup>These messages were sent to all patients recorded as drinking alcohol at risky levels, including those who were also recorded as smoking. These data were merged owing to the low number of patients recorded as drinking at risky levels.

<sup>c</sup>RR: relative risk.

<sup>d</sup>Not available.

<sup>e</sup>N/A: not applicable.

### Acceptability (After Implementation)

The survey was completed by 3.67% (56/1526) of participants. Among the 56 survey responders, 18 (32%) participants reported that they had clicked on the link within the SMS text message and 9 (16%) participants reported that they downloaded the app. Approximately two-thirds (36/56, 64%) of the participants found the messages to be at least slightly helpful. The message was found to be not at all difficult by almost all patients (55/56, 98%). Most participants (51/56, 91%) were happy with the wording of the SMS text message. Almost half of the participants (26/56, 46%) reported that they would like the hospital to deliver this service; however, an opt-out option was also popular (21/56, 38%). Most of the respondents (42/56, 75%) submitted free-text comments. Common positive themes were that the SMS text messages were supportive (7/56, 13%), easy to understand (6/56, 11%), and brief (6/56, 11%). Common...
negative themes were that the messages were irrelevant (9/56, 16%), not supportive (4/56, 7%), and unwanted or unexpected (3/56, 5%).

Discussion

Principal Findings

Our electronic signposting strategy that used SMS text messages to promote health behavior change apps to patients at risk via an acute care hospital EHR was found to be acceptable and feasible and to achieve clinically meaningful reach. The hospital sent 1526 SMS text messages signposting patients to either the NHS Smokefree or Drink Free Days apps, with 13.56% (207/1526) of the patients clicking on the embedded link to the apps. The level of reach exceeded our 5% a priori success criterion in a non–help-seeking population of patients receiving unsolicited SMS text messages, disconnected from hospital visits. The strategy was found to be acceptable both before and after implementation. The OptiMine study demonstrated the proof of concept for this novel implementation strategy, specifically signposting digital health promotion interventions at scale to at-risk individuals who are not seeking help.

A 13.56% (207/1526) click rate is promising, considering the comparable outcomes reported in the literature. For example, in a study to promote colorectal cancer screening, patients—identified from EHRs as being overdue for their colorectal screening—received an electronic message from their primary care physicians with information about their overdue status, methods to arrange a screening appointment, and a web-based risk assessment tool for colorectal cancer [33]. The intervention led to 3% of the patients requesting colorectal screening. Our 13.56% (207/1526) click rate exceeds this result. However, it is important to account for the action requested from patients, which will trigger varying levels of response, such as the 3% response rate for colorectal screenings versus the 13.56% (207/1526) click rate on our embedded link to behavioral interventions. Future research should build on this study and aim to define appropriate thresholds for e-referral interventions that correspond to different behaviors that require varying degrees of involvement from patients (eg, time and effort invested, duration of behavior, and difficulty of behavior). Furthermore, it is important to note that our SMS text messages were sent from the patients’ hospital, which would be expected to foster trust in the messages and their source and increase uptake.

Using SMS text messages to signpost patients to addictive behavior apps is a scalable implementation strategy, which can provide individualized support to patients at risk. As such, cost-effectiveness increases as the strategy is scaled up. Our low-cost and low-burden strategy has the potential to reach large numbers of patients at high risk in a novel form (proactive referral to existing tools). Health promotion is typically delivered in primary care and community settings, but it is equally important in acute care settings, as health behaviors such as smoking and risky drinking can cause long-term health conditions and exacerbate existing health conditions. Following the success of this study, the WSFT plans to routinely deliver health promotion advice using electronic signposting.

Strengths and Limitations

The mixed methods design helped to explore multiple implementation outcomes that assess both proximal outcomes and indicators of success [23]. This was a novel and pragmatic study, facilitated by NHS staff in a busy hospital setting. Major strengths of this study included the real-world application of the strategy with patients at the hospital and readiness of the implementation context. Senior-level leadership and buy-in from a public health consultant within the digital health team were instrumental in the successful implementation of this project. Global Digital Exemplar Trusts have EHRs and in-house expertise to support the setup and delivery of SMS text message signposting. We identified the following stakeholders as pivotal to the successful setup, implementation, and evaluation of the strategy: patient representatives, information analysts (assess infrastructure and access to data), deputy chief information officer (design and oversight of message implementation), IT integration developer (implement SMS text message signposting and send messages), public health manager (in-house data analysis), communications team, volunteer coordinator, and administrative support. Although the readiness of the implementation context is considered a strength of our study, it could be seen as a limitation to broad scalability in less-ready, low-resourced contexts, which may serve more disadvantaged or underserved populations.

There were challenges in using the EHR as a research database. The data set was extracted directly from the EHR at different time points owing to the size of the EHR, time to download the data, and workload of hospital staff. Queries for different parts of the data were run on different days over a 4-month period from October 2019. Each day, the total number of EHR records changes owing to new records being created, existing records being edited, and people dying. Smoking and alcohol consumption status were not retrievable from hospital day patients, which meant that there were large amounts of missing data. Furthermore, there were difficulties in retrieving accurate health condition data, and matching to eligible patients was not possible. Health condition data rely on the population of the appropriate fields in a patient record; if the condition is listed in the wrong place or not recorded at all, it cannot be electronically retrieved. This occurs with some frequency across the EHR, and this limitation needs to be addressed before using electronically retrieved health data for similar studies in the future. As such, the full capacity of patient reach was not used owing to missing data, and the impact of long-term health conditions on the likelihood of clicking the embedded link is unknown.

Our primary outcome—whether each participant did or did not click the embedded link—directly measured the willingness to access information about behavior change apps among a non–help-seeking population. However, we did not measure the proportion of participants who actually downloaded and used the app. Although technical limitations prevented us from measuring app use, uptake should be assessed through future studies to support broader dissemination of this approach. In addition, response rates to the web-based questionnaire were low and may have been subject to response bias. This may have been owing to message fatigue, where patients had previously...
received up to 2 SMS text messages signposting to the apps, followed by 2 messages inviting participation in the survey. Finally, the SMS text messages did not include instructions for opting out of future messages, which is a requirement in the United States. Future studies should examine click rates in the presence of explicit opt-out instructions.

Implications for Further Research and Practice
The WSFT views this approach as the future for routine delivery of digital interventions for health promotion within their acute care context and as an important adjunct to the opportunistic behavior change interventions that are made by health care professionals. Further refinement and evaluation are needed to optimize the SMS text message content and delivery approach. Additional studies with other populations are needed to understand the best strategy for implementation at other hospitals and institutions. The COVID-19 pandemic has highlighted the urgent need for scalable health promotion support, which can be delivered via digital technology. There is also an international call for greater access to high-quality, safe, and effective addictive behavior apps [34]. Regarding refinement and further evaluation of SMS text message signposting to apps, we propose the following ideas for further development:

1. Tailor message content to groups of people who are less likely to engage, such as older people, low socioeconomic status groups, and ethnic minority groups.

2. A timely trigger for the SMS text message, possibly in the context of a face-to-face consultation, may be more effective than sending all the messages at the same time. The SMS text messages were sent in January 2020 to optimize the click rate by taking advantage of the seasonal high demand for support, and therefore, other times of the year may be less or more likely to have high rates of engagement.

3. Use the strategy to target other health behaviors.

4. Explore other implementation outcomes, such as implementation cost and cost-effectiveness, compared with other methods for promoting lifestyle change; sustainability; and fidelity, including engagement with the app.

5. Investigate system interventions for addressing barriers identified regarding the reliability of EHR data.

Conclusions
The OptiMine study demonstrated the proof of concept for this novel implementation strategy, which used SMS text messages to signpost at-risk individuals to behavior change apps at scale. The level of reach exceeded our a priori success criterion for a non–help-seeking population of patients receiving unsolicited SMS text messages, disconnected from hospital visits. These findings suggest that electronic signposting to support is an effective method for health systems to proactively engage meaningful proportions of their at-risk populations.

Acknowledgments
The authors are grateful to the patients and staff at the West Suffolk Foundation Trust, who participated in this study, and the team that facilitated the study on-site, namely, Adam Jull, Cassia Nice, Josh Wigley, Liam McLaughlin, Simon Reeve, Anna Hollis, Penny Molkenthin, and Paul Oats. This study was funded by a Cancer Research UK Cancer Policy Research Centre innovation grant (C69129/A28665). The authors acknowledge the support from the National Institute for Health Research (NIHR) Clinical Research Network. The effort of SE was supported by the Division of Intramural Research of the National Institute on Minority Health and Health Disparities, National Institutes of Health. ATR was supported by the National Institute on Drug Abuse grant (K12DA041449). This is a summary of the study supported by the NIHR Applied Research Collaboration East of England. The views expressed are those of the authors and not necessarily those of the National Health Service, NIHR, UK Department of Health and Social Care, US National Institutes of Health, US Department of Health and Human Services, or West Suffolk Foundation Trust.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Qualitative focus group findings.

References


14. Khadjesari et alJMIR FORMATIVE RESEARCH


Abbreviations

AUDIT-C: Alcohol Use Disorders Identification Test–Consumption
CONSORT: Consolidated Standards of Reporting Trials
EHR: electronic health record
NHS: National Health Service
NIHR: National Institute for Health Research
RR: relative risk
WSFT: West Suffolk NHS Foundation Trust

Edited by A Mavragani; submitted 14.10.21; peer-reviewed by M Bird, W Cha; comments to author 08.01.22; revised version received 12.03.22; accepted 31.03.22; published 11.07.22.

Please cite as:
Novel Implementation Strategy to Electronically Screen and Signpost Patients to Health Behavior Apps: Mixed Methods Implementation Study (OptiMine Study)
JMIR Form Res 2022;6(7):e34271
URL: https://formative.jmir.org/2022/7/e34271
doi: 10.2196/34271
PMID:
A Web Platform for Standardized Data Acquisition, Processing, and Export in the Child Psychopathology Clinical Routine (MedicalBIT): Design and Implementation Study

Paola Colombo¹, MSc, PhD; Silvia Busti Ceccarelli¹, MSc; Stefano Pacchiarini², ING; Stefano Cribellati², ING; Massimo Molteni¹, MD

¹Child Psychopathology Unit, Istituto di Ricovero e Cura a Carattere Scientifico Eugenio Medea, Bosisio Parini, Italy
²SEGE Srl, Milan, Italy

Corresponding Author:
Silvia Busti Ceccarelli, MSc
Child Psychopathology Unit
Istituto di Ricovero e Cura a Carattere Scientifico Eugenio Medea
Via Don Luigi Monza 20
Bosisio Parini, 23842
Italy
Phone: 39 +39 031 877623
Email: silvia.busti@lanostrafamiglia.it

Abstract

Background: The rapid extent of digital innovation for the collection of data has transformed the way in which health professionals collect, share, and analyze health information for better clinical decision-making and health care. In the last decade, there has been an increased interest in telemedicine by mental health agencies; the gap between the need for care and both diagnosis and treatment is wide, and digital technology could play an important role in filling this gap. However, there are limited data on the effectiveness of the clinical process and cost-effectiveness of most telemedicine applications.

Objective: This study examined the implementation of the first Italian online, web-based, comprehensive screening tool and described the screening and diagnostic process through the interactive web platform in a child psychopathology clinic. This is a feasibility study that aims to present the design and implementation of the best practices to improve patient experiences and clinical outcomes. Moreover, the paper evaluates the platform with qualitative and quantitative measures.

Methods: We planned, designed, and implemented a web-based system to collect, store, and manage clinical data. The platform was developed by a multidisciplinary team composed of researchers, clinicians, and informatics professionals through different steps. First, we defined the clinical information to be collected. A number of measures were chosen, tapping several clinical risk areas such as neurodevelopmental disorders and emotional and behavioral problems. The web application architecture and process were then designed. The three phases of process design are described in detail: design of the input interface, processing design, and design of the output interface. Finally, the system has been implemented and evaluated. Based on indicators recommended by the National Quality Forum and the Italian National Guidelines, we evaluated the quality of the system and used quantitative measures that were replicable and comparable over time.

Results: We present the implemented architecture and features of Medea Information and Clinical Assessment On-Line (MedicalBIT), and we provide performance measures for the data collected between October 2018 and June 2021. The measured concepts pertain to four domains: access to care, financial impact/cost, experience, and effectiveness.

Conclusions: In this study, we present the successful implementation of an innovative digital tool. The findings of this study show that the implemented web-based platform appears to be an efficient, cost-effective, and feasible way to improve digital care in the field of child psychiatry.

(JMIR Form Res 2022;6(7):e36757) doi:10.2196/36757

KEYWORDS
digital health; big data; developmental psychopathology; neurodevelopmental disorders; digital data; digital innovation; mental health; screening tool; children; psychopathology; web platform; digital intervention; clinical outcome
Introduction

Digital data collection is an emerging trend in various fields, including the medical and psychological ones. Digital innovation is changing the way health information is collected, shared, and analyzed for better clinical decision-making and health care. Digital innovation is also rapidly expanding in the medical field, significantly improving the quality of health care, reducing health care costs, and enhancing research processes [1]. The rapid evolution of technology has recently promoted web platforms for big data collection in the health care field, and related publications are exponentially increasing [2].

“Big data” refers to a voluminous collection of information taking place quickly without affecting quality, and there are several well-known definitions describing this [3-5]. Pastorino and colleagues [2] stress the importance of defining data as both smart and big because big data presents a substantial potential when it is meaningful. This implies using data for improving health conditions by searching for increasingly clearer and more accurate links between causes, diseases, therapies, and outcomes. In the Study on Big Data in Public Health, Telemedicine, and Healthcare [6], the European Commission identified four macroareas in health care for big data use: (1) early signs for detection, diagnosis, and intervention; (2) identification of risk factors for diseases to improve prevention; (3) enhancing pharmacovigilance and patient safety by communication of real-time information; and (4) improvement in outcome prediction.

All this is possible since a large amount of data is an invaluable resource for epidemiological studies, analyses of general population needs, treatment evaluation, and experimental designs on the target population. If research develops in this direction, it will enable precision medicine that will contribute to optimization of resources: the right care for the right patient at the right time. Therefore, smart use of big data can provide a possible answer to the need for care: socioeconomic and clinical sustainability. In line with this, health technology assessments aim to inform on safe, effective, patient-centered and digital technology could play a fundamental role [12]. In the child and adolescence mental health services, the parents and caregivers of young patients must be involved in the clinic diagnostic process, both as powerful sources of information to be used by clinicians and to obtain a clear understanding about their child’s difficulties. This engagement should be conducted in a comfortable environment for the participants.

Telemedicine seems to be efficient both for the assessment and treatment. For example, among services for the diagnosis of autism spectrum disorder (ASD), the use of telemedicine has shown some promising results in terms of observation [13-15]. As for remote treatment, several recent studies on telemedicine and developmental disorders such as autistic spectrum disorder, specific learning disorder, specific language impairment, dyspraxia, and acquired brain injuries show encouraging results [16-20].

A review of the literature by the National Quality Forum—a no-profit organization for health care development—found a positive effect of telehealth on the quality of processes, outcomes, and costs. It also identified the medical areas where telehealth spread more extensively, with a correspondingly greater increase in scientific publications, namely, dermatology, mental health, rehabilitation, medical management, and chronic diseases. However, despite this topic being current and debated, telemedicine is relatively recent, and there is a lack of studies assessing its quality [21].

In our study, we examined the implementation of the first Italian online, web-based, comprehensive screening tool: Medea Information and Clinical Assessment On-Line (MedicalBIT) [22]. We carried out a feasibility study aimed to present the design and the implementation of the best practices to improve users’ experiences and discuss the evaluation of the platform with qualitative and quantitative measures. Additionally, we intend to describe a web-based screening and diagnostic process in a child psychopathology clinic.

Methods

Overview

This section describes the key elements of the platform implementation.

We planned, designed, and implemented a web-based system to collect, store, and manage clinical data through different steps. The whole eHealth system is composed of users (patients and clinicians), a service provider (the Association La Nostra Famiglia-IRCCS Eugenio Medea, a no-profit organization providing care and rehabilitation to children with disabilities), and a service developer (SE-GE Consulting Company).

The main purpose of the MedicalBIT platform is to support diagnostic flow by systematic data collection.

https://formative.jmir.org/2022/7/e36757

JMIR Form Res 2022 | vol. 6 | iss. 7 | e36757 | p.196

(page number not for citation purposes)
Patients can answer questionnaires comfortably from home through a user-friendly and easily accessible interface, saving time traveling to and staying in a clinic. Clinicians have access to a real-time description of individual patients’ symptoms and a graphical output of possible alerts through clinically based algorithms. Clinicians then assign patients to different diagnostic paths for principal neurodevelopmental disorders (Attention-Deficit/Hyperactivity Disorder, Autism Spectrum Disorder, Specific Learning Disorders, Language Disorders) and behavioral and emotional disorders based on this output. Taking into account the international gold standard, a clinical assessment is performed for each path. The final diagnostic output may differ from the initial path assignment based on the assessment results.

Finally, the data collected can be easily exported in Excel (Microsoft Corporation) format for research analyses: researchers can use collected data for further analysis to develop predictive models.

The platform was developed by a multidisciplinary team composed of researchers, clinicians, and informatics professionals. Figure 1 provides an overview of the workflow, and Figure 2 displays some picture frames of the web platform home page. A description of each step is reported below:

1. Defining clinical information to be collected
2. Web application architecture and process design
3. System implementation and evaluation

**Figure 1.** The diagnostic flow on Medea Information and Clinical Assessment On-Line (MedicalBIT).
Defining Clinical Information to Be Collected

A number of measures were chosen by tapping several clinical risk areas, such as neurodevelopmental disorders and emotional and behavioral problems. Measures were selected for their feasibility in routine clinical practice (ie, brevity, free availability, validation in children and young people, and translation) and psychometric performance (ie, validity, reliability, and sensitivity to change).

The following questionnaires were selected:

- **Risk factors**: A questionnaire to explore biological and environmental risk factors such as family composition; presence/absence of psychiatric diseases in parents or close relatives; prenatal, perinatal, and postnatal factors; and developmental milestones.
- **ASD**: The Modified Checklist for Autism in Toddlers [23] and the Autism Spectrum Quotient: Children’s Version [24] were used for the detection of ASD. The first is one of the most widely used ASD toddler screening instruments; it is easily accessible and low-cost. The second is a brief, parent-reported, 50-item questionnaire to quantify autistic traits in children aged 4-11 years.
- **Emotional and behavioral problems**: The Strengths and Difficulties Questionnaire [25] is a brief instrument widely used to assess main areas of developmental psychopathology and personal strengths. It consists of 25 items and is available in three forms depending on responders: parents, teachers, and adolescents (self-report).
- **Other neurodevelopmental disorders**: Ad hoc screening tools for specific language impairment and specific learning disorder were implemented by our institute’s research group working on specific learning, language, and communication disorders.

Overall, these tools cover ages ranging from 18 months to 15 years.

After this initial screening step, other standardized assessment tools can be administered too, such as the Child Behavior Checklist (CBCL) [26,27] and the Development and Well-Being Assessment (DAWBA) [28]. Through an interoperability system, links and credentials can be sent to caregivers to complete questionnaires (CBCL thorough ASEBA-Web) and the interview (DAWBA thorough Dawba.net), and results can then be imported by uploading appropriately encrypted files.

Web Application Architecture and Process Design

Figure 3 shows a high-level diagram of the platform development process. The three phases of process design are described in detail below:

- Design of the input interface
- Processing design
- Design of the output interface
**Design of the Input Interface**

An easy-to-use interface was developed to support the interaction between front-end users and the web-based application; the web interface is simple to learn and easy to use, and it displays information in a consistent and progressive manner and maximizes functionalities. To be displayed properly on most mobile devices, the platform implements a responsive web design.

The system also relies on the active engagement of caregivers who send their data to the informatics system using a PC and mobile device. Patients scheduled for an appointment at the Child Psychopathology Unit (IRCCS Medea) get an email with a token for the first web app. They are then guided through a series of steps: completion of a brief form with primary clinical information, verifying of this information, and filling out of the registration form. They then receive a set of questionnaires according to their age and main characteristics, along with the link to the MedicalBIT assessment for completion.

Users are not asked to complete all forms in one session, they can resume from where they last left. After completing all the selected questionnaires, patients are scheduled for a doctor’s visit.

To support the user’s interaction with the platform, a user-friendly back-end interface was implemented. Through a control panel, staff can handle main functions such as administration rights, patients’ lists, and support for data entry.

**Processing Design: Technical Implementation of Data Storage and Automated Scoring Algorithm**

At the end of each survey, the system immediately processes the caregiver’s answers using artificial intelligence algorithms. Starting from scores provided for every possible answer and relying on appropriate threshold values, alerts, and key performance indicators are generated, which are useful to health professionals not only to monitor single patients but also for statistics and research purposes.

The whole process is General Data Protection Regulation compliant. The web portal is protected with a digital certificate issued by a certification authority, implements HTTPS (the secure browsing protocol for the World Wide Web), and is hosted on a virtual server in the GARR cloud environment. GARR is the Italian ultrabroadband network for education and research.

The infrastructure provides high reliability, service continuity, and data protection. Backup takes place daily, and we rely on disaster recovery as a service in an alternative environment to ensure availability in case of downtime. We used open source solutions for easier integration with company applications and to avoid vendor lock-in. Our platform is compliant with Italian guidelines [29]: “To ensure an effective support to all company processes, it is necessary to guarantee data univocity and integrity, real-time updating, historicization and audit trail, ergonomics, standardization, integration, stability, availability, security, privacy, innovation, evolvability.”

All data are unique; nothing is duplicated. Information can be accessed only by authorized users and is protected from unauthorized changes; furthermore, confidential data is safeguarded.

There are two different access levels:

- **Front-end user**
- Caregivers filling out the forms can only view and download the information concerning their child. They have no access to scores generated by the automatic
scoring algorithms. Multiple patients can be associated to each caregiver (eg, in the case of siblings).

- Back-end users
  - Health care professionals viewing the answers and any alerts enter the diagnoses at the end of the diagnostic process.
  - Back-office users with full operation rights
  - The system administrator has access to additional features such as dashboard and data export (see next paragraph).

The integrity of data and documents is guaranteed by the use of appropriate logs of activity and changes. These logs are also useful from a legal point of view. All information relating to a particular patient can also be deleted by users with administrator rights. Documents saved on the server are protected by encryption; to comply with the privacy legislation, the information stored in the database is pseudo-anonymized.

**Design of the Output Interface**

A real-time graphing, panel, and table interface is available only to back-end users and contains all processed data in real time. This data is also displayed in interactive dashboards available at different access levels. Two different types of reporting are created to display the collected data (Figure 4):

1. Clinical output: an individual-level output, reporting patient outcome scores
2. Dashboard: group-level data, reporting summarized data in a dashboard

All data stored in the database system can be exported in Excel format for further processing.

**System Implementation and Evaluation**

This digital system has been on since October 2018 and has been used by 1301 patients (front-end users) and back-end users, including 17 clinicians, 2 back-office operators, and administration operators.

Back-end users are employed by the provider service: back-office operators are office secretaries and administration operators are psychologists and researchers who designed the service in collaboration with the data scientist and consulting company SEGE srl.

Based on indicators recommended by the National Quality Forum [21] and the Italian National Guidelines [29], we evaluated the quality of the system and used quantitative measures that were replicable and comparable over time. As shown in Textbox 1, the selected measures pertain to four areas: access to care, financial impact/cost, experience, and effectiveness.
**Textbox 1.** Telehealth measurement framework domains and subdomains.

<table>
<thead>
<tr>
<th>Access to care</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Access for parent/caregiver</td>
</tr>
<tr>
<td>• Access to information</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>• System effectiveness</td>
</tr>
<tr>
<td>• Operational effectiveness</td>
</tr>
<tr>
<td>• Technical effectiveness</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The financial impact to family/caregiver</td>
</tr>
<tr>
<td>• The financial impact to care team</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Patient, family, or caregiver experience</td>
</tr>
</tbody>
</table>

**Results**

According to the Italian National Guidelines [29] and the National Quality Forum [21], we provide performance measures available for data collected between October 2018 and June 2021 (Table 1 and Figure 5). This process is a work in progress, and in the future, we aim to capture more indicators about different domains and subdomains.
<table>
<thead>
<tr>
<th>Domains and subdomains</th>
<th>Qualitative measures</th>
<th>Quantitative measures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Access to care</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Access for parent/caregiver</td>
<td>• The web system is responsive to different devices</td>
<td>• N/A&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>• Access to information</td>
<td>• Family’s patient can access the system using dedicated access points in the institute</td>
<td></td>
</tr>
<tr>
<td>•</td>
<td>• Clear instructions in the home page</td>
<td></td>
</tr>
<tr>
<td>•</td>
<td>• Responsive technical assistance</td>
<td></td>
</tr>
<tr>
<td><strong>Effectiveness</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• System effectiveness</td>
<td>• N/A</td>
<td>• Continuity: from October 2018 to June 2021</td>
</tr>
<tr>
<td>•</td>
<td></td>
<td>• Dimension: 1100 new patients accessed the telehealth system and were subsequently sorted to different diagnostic paths. For details about the accesses trend over time, see indicators in Figure 5. Each patient completed on average 4 questionnaires, for a total of 177 items.</td>
</tr>
<tr>
<td>•</td>
<td></td>
<td>• Speed: (1) Only a few days elapsed between the initial request from family and the completion of screening questionnaires on the platform. (2) The average length between screening questionnaire completion and completion of the diagnostic process was 106 days, with a decrease over the 3-year period from 2019 to 2021 (see Figure 5e)</td>
</tr>
<tr>
<td>• Operational effectiveness</td>
<td>• The system is perfectly integrated within the traditional diagnostic path.</td>
<td>• N/A</td>
</tr>
<tr>
<td>•</td>
<td>• Customized in-person visits are based on web questionnaire output.</td>
<td></td>
</tr>
<tr>
<td>• Technical effectiveness</td>
<td>• Output dashboard for clinicians and database with collected data for researchers are automatically generated.</td>
<td>• N/A</td>
</tr>
<tr>
<td>•</td>
<td>• The system is integrated with other screening and diagnostic tools implemented on different platforms: the Development and Well-Being Assessment [28] and the Child Behavior Checklist [26,27].</td>
<td></td>
</tr>
<tr>
<td><strong>Costs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Financial impact for family/caregiver</td>
<td>• Decrease in the length/frequency of stay/visit to hospital</td>
<td>• N/A</td>
</tr>
<tr>
<td>• Financial impact for care team</td>
<td>• Clinician can integrate traditional work setting with smart working</td>
<td></td>
</tr>
<tr>
<td><strong>Experience</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Patient, family, or caregiver experience</td>
<td>• N/A</td>
<td>• Caregivers dropout: (1) 5.3% of registered users do not fill in any questionnaire. (2) 99.6% of caregivers that begin a questionnaire, complete it</td>
</tr>
</tbody>
</table>

<sup>a</sup>N/A: not applicable.
**Discussion**

### Principal Results

In this study, we present the design and implementation of the MedicalBIT web platform, which allows patients to complete questionnaires comfortably from home through a user-friendly and easily accessible interface.

This work is based on the European Commission’s recommendations in the Study on Big Data in Public Health, Telemedicine, and Healthcare [6] that identified four major areas in health care where big data may be used: (1) early signs for detection, diagnosis, and intervention; (2) identification of risk factors for disease to improve prevention care; (3) enhancing pharmacovigilance and patient safety by communication of real-time information; and (4) improvement in outcomes prediction.

The MedicalBIT platform, as described in this paper, allows for the timely collection of clinical data to support the diagnostic process and subsequent steps (1 and 3), and it enables the development of predictive models to improve preventive care and outcome prediction (2 and 4). A core component of the proposed model is caregivers’ perspective of their child’s mental health, laying the foundation for evidence-based practice and person-centered care. The MedicalBIT web platform was introduced into the clinical workflow, making patients’ and caregivers’ care experiences more comfortable. As mentioned, health information technology can save time and reduce costs, thereby improving the health care experience for both patients and clinicians. For this purpose, we also collected customer satisfaction feedback.

Furthermore, performance measures enable us to describe the platform with an objective measurement that allows for comparisons both within the same platform over time and between different web platforms used for similar purposes.
According to reports so far, available quantitative data shows general upward trends for platform use from 2018 to 2021 despite the apparent drop during the first months of the COVID-19 pandemic, which required a complete reorganization of care provision.

This trend is in line with research confirming the value of telepsychiatry [30,31] not only during the pandemic period but also before and, most importantly, after the emergency. As described in the Introduction, use of telemedicine in the psychiatry and developmental psychiatry fields can offer several benefits such as improved efficiency and effectiveness of mental health services: it can reach more people with fewer resources [32].

We hope that these changes will lead to filling the care gap in the field of child psychiatry, which has been appropriately defined as “one of the most difficult and crucial challenges of the next decade” [12].

Our project fits into this still uncertain and experimental scenario, with its pros and cons. Telemedicine in the child psychiatry clinical routine can help both patients and health care providers save time and provide them easy access to the care system. Furthermore, because of its fast and advantageous features, it facilitates the hospital and clinic workflow, as processing and automatic output of clinical data enables the collection of large amounts of data for research purposes. Nevertheless, there are some limitations to consider such as the impact of physical distance on human relationships and ethical and coroner issues [29].

Further studies are clearly necessary to establish evidence-based telepsychiatry-specific standards of care, following guidelines proposed by the American Telemedicine Association [33] and the American Academy of Child and Adolescent Psychiatry [34].

**Limitations and Further Development**

Based on the current limitations, we are working on a customer satisfaction questionnaire for both families and clinicians to collect feedback to make users feel more engaged in this process and help us as back-office operators to improve services. Moreover, we intend to improve our platform performance measurement tools to obtain more quantitative indicators about the four domains presented in Table 1.

According to Waller and Stotler [35], one of the next steps should be the implementation of a measurement and methodology evaluation of the impact of telehealth on clinical outcomes (in our case, on the diagnostic outcome as an indicator of clinical effectiveness). Our intent is to also implement a platform-based service that could follow and provide assistance to patients after diagnosis (information, psychoeducational materials, rehabilitation tools).

**Conclusions**

This study describes the successful implementation of an innovative digital tool. According to our results, the web-based platform appears to be a feasible, efficient, and cost-effective method for enhancing digital care in the field of child psychiatry. It also contributes to the collection of big and smart data, in line with European guidelines. This work shows how large data sets may enhance the accuracy and timing of diagnosis, and contribute to more effective interventions based on predictive models.

**Acknowledgments**

The authors wish to thank clinicians and back-office staff who took part in the web platform co-design and all clinicians using Medea Information and Clinical Assessment On-Line (MedicalBIT). This work has been possible due to the patients’ families, clinicians, researchers, and back-office staff who trust this innovative system.

**Authors’ Contributions**

PC, SBC, and MM conceptualized the study. SBC gathered the resources and curated the data. SBC, SP, and PC prepared the original draft. PC, SBC, SP, and SC reviewed and edited the manuscript. MM supervised the study.

**Conflicts of Interest**

None declared.

**References**

1. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. World Health Organization. 2016. URL: [https://apps.who.int/iris/bitstream/handle/10665/252183/?sequence=1](https://apps.who.int/iris/bitstream/handle/10665/252183/?sequence=1) [accessed 2022-06-27]


Abbreviations

- **ASD**: autism spectrum disorder
- **CBCL**: Child Behavior Checklist
- **DAWBA**: Development and Well-Being Assessment
- **MedicalBIT**: Medea Information and Clinical Assessment On-Line
- **MindLAMP**: Learn, Assess, Manage, and Prevent

©Paola Colombo, Silvia Busti Ceccarelli, Stefano Pacchiarini, Stefano Cribellati, Massimo Molteni. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.07.2022. 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.
Participant Experiences of a COVID-19 Virtual Clinical Study Using the Current Health Remote Monitoring Platform: Case Study and Qualitative Analysis

Juliana Pugmire¹, MPH, DrPH; Jessie Lever Taylor², BA, MSc; Matt Wilkes¹, MBChB, PhD; Adam Wolfberg², MPH, MD; Nicole Zahradka², PhD

¹Current Health Ltd, Edinburgh, United Kingdom
²Current Health Inc, Boston, MA, United States

Corresponding Author:
Juliana Pugmire, MPH, DrPH
Current Health Ltd
Suite 2, Ground Floor, Orchard Brae House
30 Queensferry Road
Edinburgh, EH4 2HS
United Kingdom
Phone: 44 131 285 8101
Email: juliana.pugmire@currenthealth.com

Abstract

Background: During the COVID-19 pandemic, individuals with a positive viral test were enrolled in a study, within 48 hours, to remotely monitor their vital signs to characterize disease progression and recovery. A virtual trial design was adopted to reduce risks to participants and the research community in a study titled Risk Stratification and Early Alerting Regarding COVID-19 Hospitalization (RiskSEARCH). The Food and Drug Administration–cleared Current Health platform with a wearable device is a continuous remote patient monitoring technology that supports hospital-at-home care and is used as a data collection tool. Enrolled participants wore the Current Health wearable device continuously for up to 30 days and took a daily symptom survey via a tablet that was provided. A qualitative substudy was conducted in parallel to better understand virtual trial implementation, including barriers and facilitators for participants.

Objective: This study aimed to understand the barriers and facilitators of the user experience of interacting with a virtual care platform and research team, while participating in a fully virtual study using qualitative and quantitative data.

Methods: Semistructured interviews were conducted to understand participants’ experience of participating in a virtual study during a global pandemic. The schedule included their experience of enrollment and their interactions with equipment and study staff. A total of 3 RiskSEARCH participants were interviewed over telephone, and transcriptions were inductively coded and analyzed using thematic analysis. Themes were mapped onto the Theoretical Domains Framework (TDF) to identify and describe the factors that influenced study adherence. Quantitative metrics, including adherence to wearable and scheduled tasks collected as part of the RiskSEARCH main study, were paired with the interviews to present an overall picture of participation.

Results: All participants exceeded our definition of a fully adherent participant and reported that participation was feasible and had a low burden. The symptoms progressively resolved during the trial. Inductive thematic analysis identified 13 main themes from the interview data, which were deductively mapped onto 11 of the 14 TDF domains, highlighting barriers and facilitators for each.

Conclusions: Participants in the RiskSEARCH substudy showed high levels of adherence and engagement throughout participation. Although participants experienced some challenges in setting up and maintaining the Current Health kit (eg, charging devices), they reported feeling that the requirements of participation were both reasonable and realistic. We demonstrated that the TDF can be used for inductive thematic analysis. We anticipate expanding this work in future virtual studies and trials to identify barriers and enabling factors for implementation.

(JMIR Form Res 2022;6(7):e37567) doi:10.2196/37567
KEYWORDS
virtual trial designs; virtual enrollment; digitalized health; theoretical domains framework; thematic analysis; remote patient monitoring

Introduction

Background
With the onset of the COVID-19 pandemic in 2020, we have seen rapid shifts in the way people work, engage with education and health care, and conduct their activities of daily life [1]. Many traditional clinical trials have slowed to a halt because of health care shortages and fear of increasing viral transmission [2]. Studies involving human participants have adapted to better use digitalized, decentralized, or virtual trial designs by the end of 2020 (though perhaps not as drastically as expected) [3]. Similar to remote working, virtual trial designs were a possibility that existed before the pandemic but have become a necessity for many researchers wanting to reduce the risk of transmission in human participants and the research community alike, while still conducting research [4,5].

Virtual clinical trials (VCTs) are site-less and rely on technologies such as apps, web-based platforms, wearable devices, and remote monitoring [6]. Digitized clinical trials also use technology to recruit and retain participants and for data collection and analysis [7]. Digitized clinical trials or VCTs leverage digital health technologies to improve participant access and engagement [7-9]. These trial designs have the potential to lower the cost of these studies and expand participation by making trials more accessible to participants [9,10].

With the shift to virtual, digitalized clinical trial designs, it may be helpful for study participants to understand specific implementation issues, including barriers and facilitators. Recruitment and retention in clinical trials are persistent challenges, whether traditional or virtual [7,11,12]. In VCTs, the study participant will likely have to interact with technology they may not have previous experience with, such as a remote continuous monitor, new apps for e-consenting and tracking, or daily surveys delivered by tablets [13,14]. There will almost certainly be a learning curve, with any instruction or assistance available also delivered remotely. Besides technical barriers, there may also be concerns about participant privacy when it comes to sharing sensitive health information [9].

Current Health and Risk Stratification and Early Alerting Regarding COVID-19 Hospitalization
Current Health (Current Health Ltd, Edinburgh, United Kingdom) is a medical technology company that creates a platform that enables continuous remote patient monitoring to support hospital at home programs and care [15]. The Food and Drug Administration–cleared Current Health kit includes a wearable device, which is a small, round disk that is attached to a band and worn on the upper arm. It monitors respiration rate, heart rate, oxygen saturation, skin temperature, and activity [15]. It can be integrated with peripheral devices, including those measuring blood pressure, axillary temperature, spirometry, weight, and continuous glucose. It also incorporates a tablet that can deliver surveys, reminders to take measurements (eg, blood pressure, weight), or a video connection to a health care provider or investigator. It requires approximately 5 minutes for a participant to set up the Current Health kit, including measuring and selecting the correct arm band size, and begin transmitting vital sign data via the secure wireless home hub. The home hub allows the Current Health platform to operate without an in-home Wi-Fi connection, thereby making the technology more inclusive.

The Current Health platform was used in the study, Risk Stratification and Early Alerting Regarding COVID-19 Hospitalization (RiskSEARCH; NCT04709068) [16], funded by the US Department of Health and Human Services branch, the Biomedical Advanced Research and Development Authority. Its purpose was to remotely monitor individuals who tested positive for COVID-19 infection, within the previous 48 hours, to learn more about disease progression and recovery. The enrolled participants wore the Current Health wearable device continuously for up to 30 days. Health data were collected to develop predictive models for the risks of hospitalization and death.

As part of the main study, the research team designed a qualitative substudy run in parallel to gain an in-depth understanding of the participant’s experience of taking part in a virtual study. Participants first had to show that they were eligible for the study by answering a web-based eligibility questionnaire, chose a time to connect with a study coordinator to be consented and enrolled, and finally had to set up and use the Current Health kit, which was shipped to their home address, all without meeting the study personnel in person. Once enrolled, participants were asked to answer a daily symptom survey delivered via a tablet and wear the Current Health wearable device 24 hours a day, except when charging the device or showering, bathing, or swimming. For the substudy, participants also agreed to conduct an interview of up to 40 minutes about the experience of participating in the RiskSEARCH study and using the Current Health kit.

The RiskSEARCH study did not progress beyond the pilot phase because of the changing landscape of the COVID-19 pandemic, including vaccine development and receding waves of infection, which negatively affected recruitment [17]. However, the substudy collected in-depth data on 3 participants, presented here as a case series, and qualitative analysis applying the Theoretical Domains Framework (TDF) to better understand the participant experience.

Theoretical Domains Framework
Virtual studies such as RiskSEARCH have many components that demand behavioral adaptation to adhere to the study intervention (eg, engaging in specific ways with the Current Health platform). The TDF synthesizes 128 theoretical constructs from 33 theories into a combined theoretical framework comprising 14 domains [18]. The TDF has been used to evaluate implementation problems, understand the mechanisms of change, and design interventions. The TDF helps

https://formative.jmir.org/2022/7/e37567

JMIR Form Res 2022 | vol. 6 | iss. 7 | e37567 | p.208
(page number not for citation purposes)
researchers identify and describe the factors that influence a set of behaviors (eg, study adherence). More specifically, it can help investigate implementation issues, including barriers and facilitators, to participating in studies such as the RiskSEARCH study and adopting the behavior changes necessary for adherence.

This is an exploratory piece of research based on a virtually delivered study run during the global COVID-19 pandemic from March 2021 to May 2021. The study team conducted this research to explore the participant experience for improving (1) recruitment and retention in future studies, (2) user experience with the Current Health platform, and (3) the ease with which the platform can be harnessed in other clinical studies, and in particular, virtual studies. We hope that these findings will aid other investigators to successfully conduct virtual studies and VCTs.

**Methods**

**The RiskSEARCH Main Study**

The RiskSEARCH study was a virtual, time-sensitive trial for individuals, aged >21 years, who tested positive for COVID-19 infection. The primary purpose of this study was to develop a machine learning–based algorithm to predict the likelihood of requiring a hospital stay of at least 24 hours using data collected from a remote patient monitoring wearable device and symptom surveys. This study used the Current Health platform for hospital-grade remote patient monitoring of vital signs and daily symptom surveys. Participants were recruited through advertisements on social media (Facebook, LinkedIn, etc) and word of mouth from March 2021 to May 2021. If an individual met the inclusion and exclusion criteria (Multimedia Appendix 1) and were interested in participating, they had 48 hours to enroll in the study. They were then consented and shipped a Current Health kit. The details of the main study will be published in a separate paper.

**Textbox 1.** Research objectives.

**Research objectives**

- To explore recruitment and retention for the Risk Stratification and Early Alerting Regarding COVID-19 Hospitalization (COVID-19) study
- To explore the feasibility, acceptability, and usability of the intervention, that is, the Current Health wearable device and tablet
- To explore barriers and facilitators of study compliance

**Topic Guide and Interviewing**

On the basis of the literature, our research objectives, and previous experience in developing interviews to understand engagement with digital technology, the study team designed an interview schedule (Multimedia Appendix 2) to explore barriers and facilitators around different aspects of the study and intervention (web-based enrollment, answering the daily survey, charging the wearable device, etc). One-to-one interviews were conducted by JP via telephone at a prearranged, mutually convenient time. JP was a senior clinical research scientist at Current Health at the time of this study, has >10 years of experience conducting interviews for qualitative and mixed methods research, holds a Doctor of Public Health and Master of Public Health in epidemiology, and is a woman.

**Recruitment and Procedure**

Although we planned to use a purposive sampling strategy, we changed to convenience sampling when the main study recruitment remained low. A total of 7 participants were offered the opportunity to participate in a one-to-one interview with a research team member (JP). Participants were approached by the study coordinator (JLT) through text messaging or telephone conversations after building rapport through the study enrollment process. A total of 4 participants agreed to participate in the study, and 3 interviews were conducted. A participant could not be contacted to set up the interview. No relationship was

Each day, the participants were sent a 21-question survey to complete on the Current Health tablet. The survey asked if participants experienced 8 specific symptoms (chills, fever, nausea, diarrhea, sore throat, dry cough, muscle ache, and loss of smell or taste) and whether they were better, worse, or the same as the previous day. In addition, there was a free-text response in which participants could add any other symptoms they were experiencing. Questions were also included about whether participants were likely to contact a health care provider or attend a hospital based on how they felt that day. This symptom survey was developed and piloted internally before it was shared with the RiskSEARCH study participants. Its purpose was to capture the symptoms and symptom severity associated with COVID-19 infections, to help drive the prediction model of the main study. In parallel, participants were asked to wear the Current Health wearable device for up to 30 days, taking it off only to charge (up to 30 minutes every 24 hours), shower, bathe, or swim.

**Qualitative Substudy**

We used semistructured interviews and reported the results following the consolidated criteria for reporting qualitative research checklist [19]. Specifically, we wanted to understand what it was like to use the Current Health kit and participate in a fully remote virtual study during a global pandemic. We collected in-depth data on the acceptability of the RiskSEARCH study and Current Health kit. Focused qualitative and quantitative research provided insights into the user experience of interacting with the Current Health kit, the Current Health research team, enrollment process, and participation in a fully virtual study.

**Research Objectives**

The research objectives of this study are presented in Textbox 1.
established between the interviewer and the interviewee before the commencement of the study. The participants knew that JP was a research scientist at Current Health and was interested in understanding their experience of participating in RiskSEARCH and using the Current Health kit.

Participants who agreed to participate in the interview were sent a PDF version of the informed consent form (ICF). Participants were sent the ICF via DocuSign (DocuSign, Inc) 24 hours before the interview. Participants could sign ahead of the call with the researcher or wait until the call to complete the ICF and ask any questions before signing. The researcher (JP) ensured that the participant questions were answered and that the participants understood the risks of study participation. Participants could opt out of recording the interviews, but none chose this option.

**Intervention**

Once enrolled in the main study, participants were required to wear the Current Health device at all times, except when charging the device (20-30 minutes every 24 hours) or when showering, bathing, or swimming. They were also required to keep the tablet charged and answer the daily symptom surveys delivered by the wearable device. **Textbox 2** shows the components of the intervention.

**Textbox 2. Components of the intervention.**

<table>
<thead>
<tr>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Be home to receive the Current Health kit delivered by FedEx</td>
</tr>
<tr>
<td>• Open Current Health box</td>
</tr>
<tr>
<td>• Set up home hub which includes plugging hub into the wall</td>
</tr>
<tr>
<td>• Select correct armband size using included sizing guide (out of 3 sizes)</td>
</tr>
<tr>
<td>• Charge wearable device on included dock until fully charged, indicated by green lights, and charge daily thereafter</td>
</tr>
<tr>
<td>• Insert wearable device into armband and wear next to skin under clothing</td>
</tr>
<tr>
<td>• Remove wearable device for showering, bathing, or swimming</td>
</tr>
<tr>
<td>• Charge tablet daily</td>
</tr>
<tr>
<td>• Answer daily symptom surveys delivered on the tablet</td>
</tr>
<tr>
<td>• At the completion of the main study (up to 30 days), repackage the Current Health kit back into the box and use the return label provided to arrange return</td>
</tr>
<tr>
<td>• For substudy, arrange a mutually convenient time to be interviewed</td>
</tr>
<tr>
<td>• Participate in an over-the-phone interview lasting up to 40 minutes about using the Current Health kit and participating in the Risk Stratification and Early Alerting Regarding COVID-19 Hospitalization study</td>
</tr>
</tbody>
</table>

**Data Collection**

Participant interviews were conducted over telephone and audio recorded using a laptop application (Windows Voice Recorder, Microsoft Corporation) and a handheld digital recorder as a backup. Interviews were anonymized and transcribed using Trint software (Trint Ltd) and checked, corrected, and edited for accuracy by the researcher who conducted the interviews (JP). Familiarization with the data began at this early stage. Participants were also asked to take a modified Telehealth Usability Questionnaire (TUQ) sent to them via an email link. The TUQ is a validated survey tool that quantifies the usability of telehealth implementations and services [20]. No repeat interviews were carried out, no field notes were made, transcripts were not returned to participants for correction, and participants did not provide feedback on the findings.

**Metrics**

As part of the main study, interview participants also contributed quantitative data, such as daily symptom surveys submitted via tablets. The data collected relevant to the substudy included the following variables as shown in **Textbox 3**.

The participants’ symptoms and vital sign alarms were presented alongside the qualitative results, as their clinical course may have influenced their experiences.

**Textbox 3. Data collected.**

<table>
<thead>
<tr>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Wearable adherence: the time the wearable device was worn compared with the study duration.</td>
</tr>
<tr>
<td>2. Daily survey adherence: the number of daily surveys completed compared with the number of daily surveys assigned.</td>
</tr>
<tr>
<td>3. Fully adherent, determined using 3 criteria: wearables worn for at least 20 hours a day and at least 6 days a week up to 30 days, daily survey responses at least 6 days a week up to 30 days, and a returned Current Health kit at the end of study participation.</td>
</tr>
<tr>
<td>4. Vital signs alarms: alarm thresholds were set for vital sign data going out of range, which could only be seen by the study team.</td>
</tr>
</tbody>
</table>
Analysis

A researcher (JP) conducted the interviews, transcribed the audio recordings using Trint transcription software, and coded the data using NVivo Qualitative Data Analysis Software (version 12; QSR International) [21]. We used reflexive thematic analysis [22]. Data were analyzed inductively following the steps of Braun and Clarke [22,23], specifically (1) familiarization of data, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) producing the report. Initial codes were inductively generated from the interview transcripts, iteratively condensed, and expanded into themes. The themes were then deductively mapped onto the domains of the TDF.

Ethics Approval

Ethics approval was obtained from the Institutional Review Board, Advarra (Columbia, Maryland, ethics approval number Pro00047371). The collected data were stored in compliance with the European Union General Data Protection Regulation, Current Health Research Data Management Policy, US Health Insurance Portability and Accountability Act, and Current Health Research Data Management Policy. Data were anonymized, and all personal identifiers were removed.

Results

Participant Characteristics and Quantitative Results

Overview

Participant details are provided in Table 1 and discussed in further sections.

Table 1. Participant characteristics.

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>Gender</th>
<th>Age (years), range</th>
<th>Wearable adherence (%)</th>
<th>Daily survey adherence (%)</th>
<th>Telehealth Usability Questionnaire score</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS001</td>
<td>Female</td>
<td>30 to 35</td>
<td>83</td>
<td>76</td>
<td>7</td>
</tr>
<tr>
<td>RS006</td>
<td>Female</td>
<td>40 to 45</td>
<td>63</td>
<td>90</td>
<td>_a</td>
</tr>
<tr>
<td>RS008</td>
<td>Female</td>
<td>35 to 40</td>
<td>92</td>
<td>100</td>
<td>—</td>
</tr>
</tbody>
</table>

_aParticipants did not complete the Telehealth Usability Questionnaire.

Case 1

RS001 initially reported experiencing chills, dry cough, and a sore throat. She did not report experiencing any other symptoms for the duration of her study. By day 6, RS001’s chills and dry cough resolved and did not reoccur. However, she reported a sore throat periodically throughout her 17 days in the study. (Figure 1A) Over the course of the study, the only vital sign that triggered an alarm on the Current Health dashboard was a high respiration rate, which occurred on days 0, 3, and 9.
Figure 1. Symptom survey data for participants, daily symptoms reported by participants: (A) RS001, (B) RS006, and (C) RS008. Reported symptoms varied by participants. White gaps between days indicate the participant reported feeling no symptoms. Black bars indicate the days that the participant did not complete the daily symptom survey. Red hatched line indicates the study duration ended before 30 days.

Case 2
RS006 experienced all 8 symptoms, specifically asked about in the daily symptom survey over the course of the study. Her fever resolved on day 1 and did not reoccur. Conversely, her diarrhea did not resolve until day 30. Her symptoms decreased over the 30 days of study participation (Figure 1B). This trend in self-reported symptoms aligned with her vital signs data. RS006 triggered 34 alarms, but no alarms were triggered from day 23 onward. Nearly half of the alarms triggered were for a high pulse rate with a low amount of motion detected (ie, the participant’s pulse rate was high while not exerting themselves physically). Other alarms triggered were for low oxygen saturation and a high respiration rate with a low amount of motion detected.

Case 3
RS008 experienced all 8 symptoms included in the daily symptom survey, although diarrhea was reported only once on day 16. Nausea was the most persistent symptom, continuing until day 28. No symptoms were reported on days 29 or 30. As with the other participants, RS008’s symptoms improved over the course of the study (Figure 1C), which was also reflected in her vital signs data. RS008 triggered 10 alarms during the 30-day study period. No alarms were triggered after day 19. Although all participants were asked to follow a link provided via email to complete the TUQ survey, only one participant completed the survey. This participant scored strongly agree (7 on the 7-point Likert scale) to all 21 questions of the TUQ, indicating high levels of usefulness, ease of use, effectiveness, reliability, and satisfaction with the Current Health kit.
Qualitative Results

Overview

In all, 3 interviews were conducted toward the end of study participation when participants were feeling better. Interviews ranged from 18 to 35 minutes. Inductive thematic analysis helped identify 13 main themes and subthemes associated with the participant experience of using the Current Health kit and being part of the RiskSEARCH study and included (1) Participant Situations, (2) Getting Started, (3) Study Support, (4) Study Communication, (5) Protecting and Contributing, (6) Determination, (7) Study Pros & Cons, (8) Optimism, (9) Uncertainty, (10) Payment, (11) Accessing Data, (12) Memory & Reminders, and (13) Making Habits.

These themes were deductively mapped to the TDF domains. These domains are described below and presented in Table 2. There were 3 domains of the TDF in which we did not match any data to: Intentions, Goals, and Emotion.

The main domains of the TDF, which we were able to map our themes onto, were Environmental Context and Resources; Knowledge combined with Skill; Social/Professional Role and Identity; Beliefs about Capabilities; Optimism; Beliefs about Consequences; Reinforcement; Memory, Attention, and Decision Processes; Social Influences; and Behavioral Regulation.
<table>
<thead>
<tr>
<th>TDF constructs</th>
<th>RiskSEARCH themes</th>
<th>Description</th>
</tr>
</thead>
</table>
| **Environmental context and resources** | Participant Situations: being sick with COVID-19 infection; caretaking responsibilities | • Study participants were recruited and went through the study after testing positive for COVID-19 infection  
• Study required steps that may have been more challenging for participants who had many caretaking responsibilities |
| **Environmental context and resources** | Getting Started; enrollment; kit components; for example; wearables; unknowns; suggestions | • Participants had to self-navigate through a web-based enrollment system and website  
• Current Health kit required setup by the participant themselves (though they did have access to technology and study team support)  
• Communication between tablet or wearable and participant  
• Wearable needed charging 30 minutes every 24 hours. Participant did not know the battery level of the wearable, but green lights on charger indicated that it was fully charged  
• Suggestions for improving any aspect of the Current Health kit |
| **Knowledge combined with skill** | Study Support\(^1\): personnel; materials | • Technology support was available 24/7 to help with any aspect of setting up or using the wearable  
• The study team was available on demand to answer any questions relating to the study or Current Health kit  
• The Quick Start Guide was available in a hard-copy booklet in the Current Health kit or digitally accessible via the tablet  
• “Study Support’ double coded with Social Influences” |
| **Knowledge combined with skill** | Study Communication\(^2\); passive; active | • Website as a source for information on the study and COVID-19 pandemic  
• “Study Communication’ double coded with Social Influences” |
| **Social or professional role and identity** | Protecting and Contributing | • Help or protect others; feeling a sense of community responsibility; wanting to help in a difficult time; being someone who helps  
• A sense of contributing toward the management of the COVID-19 pandemic |
| **Beliefs about capabilities** | Determination | • The level of commitment while engaging with the Current Health kit—survey or tablet or wearable |
| **Optimism** | Study Pros and Cons | • Positive and negative aspects of the study |
| **Optimism** | Optimism | • Seeing the positive in the bad situation of being tested positive for COVID-19 infection |
| **Beliefs about consequences** | Uncertainty | • Feedback regarding user’s “performance” or whether kit was working properly |
| **Reinforcement** | Payment | • Study participants expressing their views on the US $100 offered for their time and effort |
| **Reinforcement** | Accessing Data | • Having access to own data  
• The wearable device does not transmit data to the participant |
| **Memory, attention, and decision processes** | Memory and Reminders | • Remembering to charge and wear the wearable  
• Reminders to take the survey every day  
• Reminders to charge the tablet and take the survey |
| **Social influences** | Study Support\(^1\): personnel | • Possibility to contact technology support or study team  
• “Double coded with Knowledge combined with Skill” |
| **Social influences** | Study Communication\(^2\); active | • Via email, text, or telephone call  
• “Study Communication’ double coded with Knowledge combined with Skill” |
Environmental Context and Resources

Many contextual factors impacted the participants’ ability to successfully participate in the RiskSEARCH virtual study, including the Current Health kit functionality and design, being sick with COVID-19 infection, caretaking responsibilities, and comfort of the wearable device.

First, they had to navigate through a web-based enrollment process that included clicking on an advertisement that took the participant to a brief eligibility screening questionnaire and onto an appointment booker to connect with a member of the study team for consent. Although 2 participants said that the enrollment process was smooth and easy, another participant reported some minor problems that required contact with the study team:

I think it was pretty easy. [RS008]

So when I went to enroll, it didn’t give me a time slot, like it said that there was no one available. I guess you have to like, talk to someone at first and I remember it led me - it led me all the way to the end. But then I said, like, there was no - no one available for my initial call... So I emailed and then they were available. But it was like, still on the same day. I feel like it was a glitch or something. [RS001]

Second, participants were required to set up the Current Health kit that was delivered to their home, select the correct-sized armband, charge the device and tablet, insert the device into the armband, and begin wearing it. In all, 2 participants described the setup as an easy process with a participant providing negative feedback on the tablet stand, which they described as nonessential and fussy:

Because like, one of the first things in the instructions is...take out the stand, put the...um...put the tablet on the stand. Like I just said, you could take all of that out because at first I thought I had to do it for it to get started and I didn’t, and it wasn’t like standing up and it just seemed like a waste. [RS001]

It was super easy in the box set up that, you know, getting the tablet and everything and then getting the little charging dock. And I mean, it was easy and then I got it connected to my wifi and started wearing it that day. [RS006]

The third participant experienced problems during setup. The Current Health kit shipping was delayed, and when she began setting it up, there was a problem with connectivity, which she reported took her a few hours to work out:

And then once I got the stuff here, yeah I started setting it up and then either the mobile or the wifi wouldn’t work...it just didn’t want to connect the wifi or wearable device. [RS008]

Through the process of enrollment and setting up the wearables, participants were sick with COVID-19 infection, which meant that the usual barriers to joining a study may have been more difficult than usual. As a study team, we attempted to make the process as easy as possible for the participants:

It [enrollment] was super easy to me. I mean, even while, you know, sick as a dog with COVID I was still able to navigate and do it. So if I was able to, then you know, anybody could as long as you read and understand what you read, you can do it. [RS006]

Furthermore, study adherence may have been more difficult for participants who had caretaking responsibilities:

It’s hard when you’re a caretaker and you’ve got, you know, your mom with breast cancer. So you have to keep her schedule up plus her meds. Plus your schedule and your meds and then hubby and his meds and his schedule. It can get overwhelming, I guess, but it was just because I probably was out of my routine. [RS006]

Adherence could also be influenced by the comfort of the wearable device, which we asked participants to wear as close to 24/7 as possible, only removing the wearable device to shower, bathe, swim, or charge. They could switch arms but needed to wear it next to their skin, under their clothes. In all, 2 participants said that it was comfortable and did not give them any problems, even during sleep. A third participant provided suggestions for improvement:

Maybe if the band was...it could get...more air towards it, so you won’t get so much sweat under it...It’s you know, really gross with activity sometimes...so maybe just more airing. [RS008]

Knowledge Combined with Skill

The domains of Knowledge and Skills were combined. Participants had to acquire an understanding of how to use the wearable device and tablet for adherence. Participants were provided with a printed and digital version of the QuickStart Guide (QSG) in the Current Health box and on the tablet. All participants reported using the hard copy QSG, which was positioned to be very visible upon opening the box, and none of the participants were aware that the QSG was also available on the tablet:

I remember getting everything [Current Health kit] and then I just - as soon as you open the box, I mean, like literally step by step, as long as you follow that booklet inside. That’s what I did. I read it first. And then I started looking at stuff and I went back and I...
was like, OK, step 1 is this, step 2 is this. I mean, it hooked up in like literally ten minutes. [RS006]

When a participant encountered problems with setup or connection, they had access to the study team and Current Health’s 24/7 technical support to get things working:

I mean, I couldn’t connect, of course. Then finally I was like, OK...then I realized I could reach out to the email person or the person that was like head of tech things like and say, OK, it’s not working. [RS008]

An important aspect of technical support is the speed at which they respond, so that the participants do not become frustrated and no data are lost:

...they contacted me pretty fast, so...I didn’t think I was going to have a response like that. [RS008]

Participants had to develop skills to interact with the tablet to take daily symptom surveys and to ensure that the tablet and wearable device were charged and working. They also had to experiment with the device fit to ensure comfort:

It’s [survey] very, very easy to understand straight to the point, like you ask exactly what you need to know. And I love how it gets to all the symptoms, you can hit yes or no. And then it even asks you okay is it better than the day before this time or worse. I have loved that because some of my symptoms are a lot better and some of them are staying worse or getting worse and it varies daily. And so, yeah, I love that. That’s really cool. [RS006]

Now, if you’ve got the band too tight on your arm, your arm will hurt. That’s a learning process whenever you’re starting. I did that...Then I got it loosened up and it’s like perfect now. [RS006]

**Social or Professional Role and Identity**

The participants talked about their reasons for joining the study and contributing their time and effort. They were motivated to help in what few ways they felt they could, especially when it was difficult to help beyond isolation at home. Having a sense of contribution to efforts around the COVID-19 pandemic is important:

Look, I’m trying to be a good...I’m trying to be a good human. We’re trying to quarantine and stay away from people. [RS006]

...well, I’ll apply and see and help out the community and help out the hospital or where all this data was going to go to help you guys. See if it would do anything good for covid. [RS008]

**Beliefs About Capabilities**

The participant who had problems with the setup of the Current Health kit showed particular determination in working through the issues and troubleshooting until she could get it working. Although she had access to technical support, she was determined to troubleshoot initial connection problems independently:

And then once I got the stuff here [Current Health kit], I started setting it up and then either the mobile or the wi-fi wouldn’t work... I tried doing stuff on my own... it just didn’t want to connect to the wi-fi or wearable device... I was like OK. I’m not going to play with this anymore. And then stayed up, like all night cause I was like OK, I am not letting this thing beat me. I was just determined to like...I’m going to figure this out somehow and then yeah... I don’t even know what time it was, I don’t know, it was like twelve or one-o’clock in the morning when I finally got it to work. [RS008]

**Optimism**

A participant showed tremendous optimism in the face of the COVID-19 pandemic and her own personal trouble in being sick with the disease. This participant focused on what it was teaching her and helping her focus on gratitude for aspects of life that were going well:

It’s [COVID] definitely teaching us and it’s making me learn and making me more aware and I’m thankful for it if you wanna know the truth. [RS006]

Yeah, that’s...that’s a positive way to look at things, huh? [Interviewer] Yes, ma’am. I mean, well, you got to be positive. I was already down to the bottom. You know, I was already at the bottom of the COVID barrel. [RS006]

Although we did not hear that our participants expressed pessimism in relation to the COVID-19 pandemic or their involvement in the study, we asked them if there were any negative aspects to participating in the study. The participants mentioned that charging and remembering to charge the device and tablet were inconvenient tasks. A participant said that you could become tired of wearing the device, but she did not mind wearing it. Wearing the device could also pique the curiosity of strangers:

The only negative thing is people ask, what is that on your arm? [RS006]

Oh, interesting. [Interviewer]

Yes, that would be I would say the only thing, just getting questions about what is it? What are you wearing? And so I just tell them, I’m in a medical study since I had COVID and I’m just giving all my data and vitals and having stuff recorded. That’s what I tell them. [RS006]

**Beliefs About Consequences**

The Current Health platform at the time of this study did not relay any information to the participants. It took some time for participants to adjust to wearing and charging the device and trusting that the data were being transmitted appropriately. Participants did not always know if data would be lost if they left their homes. The biggest issue that came up for participants around charging the device was not knowing how much battery it had left, making it difficult to plan charging the device. The tablet was less of an issue because they could leave it on the charger and only use it once per day to take and submit their daily symptom survey. Several suggestions have been made to
make the battery life more apparent to the user, which are now being integrated into the product:

"I don’t know if it’s possible, but like if it told you if you needed to charge your device. Like I know it tells you, please recharge it but if there’s like a way it told you that it had a low battery the actual like...I have no idea if it’s possible but like, if it somehow would like, oh, your little arm band has low battery, charge it. [RS001]"

"So I guess, you know, I wasn’t sure if like, it would work if I wasn’t in my house. [RS001]"

After a power outage, a participant expressed concern over the lost data and whether the device was still transmitting the data:

"I was having an issue about the connection and I emailed [the Study Coordinator] and I was like, are you getting this [vital signs data] and she said, Yes, you’re fine, you’re good, it’s okay. Cause we had like a power outage so our Internet went out and everything. And I was like, Are you still getting this? Yes, it holds data for eight to ten hours. I said, Okay, just making sure because I thought I’d done messed it up [laughs]. [RS006]"

We found this domain to be linked with a concern that participants had about their own study adherence (ie, transmitting vital signs data) and desire to participate in the study to the best of their ability.

Reinforcement

A participant felt that the study would be more engaging if she could see some of her own data, and plans are underway to allow participants and patients to receive feedback on their data from wearing the device.

Participants were asked what they thought about the US $100 they were offered for the time and effort they took to participate in the study. Participants received this payment after the study ended and the equipment was returned to Current Health. All participants thought it was a nice gesture but said it was not the thing that motivated them to participate:

"I think it’s fine, like it didn’t...it didn’t sway me to participate or not participate...It’s just a nice added bonus. [RS001]"

"I think it’s more than fair...I mean, all you’re doing is just wearing this device. It’s not like you’re having to drive anywhere. You’re not having to go out of your way. All you have to do is hook up some equipment, wear the device and save the box. And then when you’re done, pack it all up and send it in. Woo hoo, I mean. It is not that hard. So I mean, I think it’s very fair. [RS006]"

"I think it’s like a nice thing...I didn’t do it for the money, I mean, I’m still waiting on that. [RS008]"

Memory, Attention, and Decision Processes

The participants in the RiskSEARCH study had high levels of adherence to wearing the device and taking the daily survey (see the aforementioned results). For the few times they did not answer the symptom survey or wear the device, we asked them to think about the reasons. A major contributor to not wearing the device was forgetting to put it back on after removing it for charging, bathing, or showering. For the tablet, it was forgetting to take the survey:

"Yes, I think I forgot to put it on...And then other than that, I don’t think I...I did stop it early because I went on vacation and I didn’t bring it with me [participant asked to stop participating before going on vacation as symptoms had resolved]. [RS001]"

"I think I pretty much wore it all the time. Sometimes...like whatever going to the shower or whatever, then maybe I...might’ve like left it on the charger a little bit longer. [RS008]"

Behavioral Regulation

Participants spoke about the importance of routine and habits for remaining adherent by wearing a charged device and having a charged tablet to take daily symptom surveys. On days when a participant was out of routine or when normal habits could not be completed, there was an increased risk of forgetting to do these things:

"I was literally running all day long from home like 9:00 that morning until about 5:30 yesterday afternoon is when I finally stopped and got home. And when I came in and made something to eat, I didn’t even...I got, I was out of my routine that day. And I didn’t even think about it. I didn’t even think about it until three o’clock this morning. [RS006]"

Social Influences

Participants had access to the Current Health technical support 24/7 and the study team close to 24/7. It was critical that we not lose participants or otherwise good adherence because of problems or questions that could not be responded to quickly. The participants could access technical support over telephone and the study team over telephone, email, or text. Participants spoke highly about the study’s clinical research associate who was primarily responsible for enrolling participants and helping them get set up:

"The lady that I was coordinating with was...she was super sweet, she was super informative. Anytime I had a question all I had to do with text or message, and she literally got back to me that same day. [RS006]"

We also looked for feedback about study aspects, such as communication, frequency of messaging, and the Current Health kit itself. We asked in-depth questions about their experience using the study website, study-related emails, texts, and telephone calls. Participants found the methods of communication acceptable and unobtrusive and said that they liked having several avenues available to them for communication with the study team.
Discussion

Principal Findings

To gain a more thorough understanding of participant experience in a fully remote virtual trial during a global pandemic, semistructured interviews were conducted with 3 of the 7 participants in the RiskSEARCH study. All participants met our criteria for being fully adherent to the study and reported through interviews or the TUQ that participation requirements in both the main study (up to 30 days of wearing a wearable device on the upper arm, responding to daily surveys, and communicating with the study team when necessary) and the substudy were feasible and low burden.

Having quick and easy access to support from the study team and Current Health technical support was a critical enabling factor for staying engaged [24]. Future virtual studies should ensure that this resource, a dedicated and responsive study team or technical support, is accessible to participants (Textbox 4). This may necessitate staffing across time zones or during out-of-office hours. Participants reported some barriers around setting up the Current Health kit, keeping devices charged, and remembering to take surveys but described these as minor challenges and showed high adherence while wearing the device and submitting responses to daily symptom surveys. Combining subjective (qualitative and self-report) and objective (quantitative like the number of surveys submitted and hours of vital signs data transmitted) data, the researchers assumed that high adherence was at least partly tied to ease of participation. A participant reported high levels of adherence (ie, reported not wearing the device on only 2 to 3 occasions that were a few hours long), while objectively showing 63% wearable device adherence. In reviewing the data, we believe that some data may have been lost when the participant was away from the home hub for >8 hours. Several factors may contribute to the differences in objective and subjective measures when collecting remote data such as perception, unknown technical issues, or improper use of digital technology. We found that the overall motivation for enrolling was a wish to contribute positively to the COVID-19 effort. In this small sample, we found adherence to be the easy part, whereas the key challenge for the research team was recruitment to the main study amidst the rapidly shifting landscape of the pandemic.

Textbox 4. Recommendations.

<table>
<thead>
<tr>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Provide participants quick and easy access to support from the study team or technical support for any digital health technologies used in the study.</td>
</tr>
<tr>
<td>• Collect in-depth information around factors that impact on study enrollment and adherence; for example, processes, communications with study team, advertisements, and trouble setting up or using technology.</td>
</tr>
<tr>
<td>• Identify discrepancies in subjective and objective measures of study adherence and try to understand why those might exist; for example, participant perception, unknown technical issues, or improper use of digital technology.</td>
</tr>
<tr>
<td>• Identify agreement in subjective and objective measures of study adherence to understand what is working well.</td>
</tr>
<tr>
<td>• Consider using the Theoretical Domains Framework (or similar framework) for assessing potential implementation problems in virtual trials.</td>
</tr>
</tbody>
</table>

The interview schedule (Multimedia Appendix 2) was developed with the purpose of understanding the participant’s experience of interacting with the Current Health kit and the ability and motivation to adhere to the study intervention. This interview schedule can be used in other qualitative studies looking to identify components of the study, such as digital technology and study materials, that facilitate or prevent adherence. We explored the barriers and facilitators to the web-based study enrollment process, communication with the study team, study advertisements and recruitment messaging, troubleshooting, burden of participating in the study, ease of use of the Current Health kit, and benefits and disadvantages of participating in the RiskSEARCH study. The data were then inductively coded into themes related to the TDF domains. The TDF is frequently used to develop interview schedules, with interviews coded into 14 domains of the TDF. The authors could not find research conducted as it was for this study, with an interview schedule developed independently of the TDF with themes from analyzed interview data and then mapped onto the TDF.

The TDF worked well for our qualitative data and the process of mapping inductive themes onto domains from the TDF was relatively straightforward. We chose the TDF because it was developed to make behavior change theories more accessible to implementation researchers [18]. It was revised and validated in 2005 [25] to help researchers identify and describe the barriers and facilitators that could influence behavior and thereby impact implementation. Evidence suggests that theoretically based health interventions are more successful than interventions with no such base [26]. We needed a method for theoretically assessing any potential implementation problems encountered while running the RiskSEARCH study; the TDF provided this method. We are unaware of the use of TDF in other virtual studies or VCTs.

The interview schedule, created independently of the TDF, produced themes that mapped most heavily onto the domains Environmental Context and Resources and Knowledge combined with Skills. We believe this is an indication of the critical aspects of the product itself, the built environment of the participants, and the knowledge and skill acquisition that are essential for setting up and using the Current Health kit.

There were 3 domains that we did not map any themes to: Intentions, Emotions, and Goals. Though we did not map any themes to the TDF domain of Intentions it was clear throughout each interview that participants made a conscious effort to be fully adherent by wearing the device as long and as often as possible.

https://formative.jmir.org/2022/7/e37567
possible, answering the daily symptom survey, and returning the Current Health kit once the study was over. As for emotions, we found that participants were content to wear the device and be adherent. For the one participant who had trouble setting up her Current Health kit, she did not describe feeling frustrated or annoyed but simply determined to get her Current Health kit working without assistance, although knowing that technical support was available. This was not a study designed to get participants to create and follow goals, which is why we likely did not have any themes that could be mapped to the domain Goals. In future interview schedules, we could consider targeting these domains to obtain the most complete picture of implementation using the TDF.

Limitations
The biggest limiting factor in this study was the sample size. Although “data saturation” is a term with some conceptual and methodological issues and is not a necessity in every type of qualitative research [27,28], this study would have benefited from more interviews and in particular from interviewees who were different ages, male, and not adherent or those who experienced technical challenges, troubleshooting, and had opportunities to develop strategies around device charging and remembering to complete the daily symptom survey. There is also a possibility that more interviews would have led to more themes that could have been mapped onto the 3 domains of the TDF, for which we did not have data. However, reaching saturation does not necessarily invalidate our findings [29]. Despite the low number of interviews, we believe these exploratory findings add value to identifying barriers and facilitators to adherence in virtual studies and specifically, studies that require using wearables and submitting daily digitally delivered surveys.

Future Work
We hope to expand these preliminary findings to future virtual studies and VCTs that will surely outlive the current COVID-19 pandemic [30]. As a study team, we are highly motivated to reduce the burden placed on study participants to make study adherence as easy and enjoyable as possible and to encourage a more diverse study population by removing logistical barriers to participation [31]. The findings from this exploratory research will contribute to the best-practices literature for VCTs and help improve the Current Health kit and study delivery for future research participants. We believe that there is also more to learn about motivating factors for participants willing to enroll and participate in infectious disease research. We are also developing a means of providing the participant’s own data to them to help with engagement and memory around charging and wearing the device and believe that this will affect adherence metrics and overall levels of study engagement.

As the COVID-19 pandemic stretches on and the need for VCTs continues to grow, there is also a need for continuing research that helps us understand the participants’ experience of engaging in these studies, including the barriers and enabling factors that influence adherence. The RiskSEARCH study did not progress beyond the pilot study because of limited recruitment, highlighting an ongoing need to improve recruitment for clinical studies, whether virtual or in person. Despite the problems with recruiting, we believe we have learned some critical lessons about the conduct of virtual study or trial presented in this paper and have produced tools to continue collecting this type of data.

Conclusions
Participants in the RiskSEARCH substudy showed high levels of adherence and engagement throughout their participation. Although participants experienced some challenges in setting up and maintaining the Current Health kit (eg, charging devices), they reported feeling that the requirements of participation were both reasonable and realistic. We have shown that the TDF can be used for inductive thematic analysis. We anticipate expanding this work in future virtual studies and trials to identify barriers and enabling factors for implementation.

Acknowledgments
This work was supported by Current Health, Ltd. Additional time on the Risk Stratification and Early Alerting Regarding COVID-19 Hospitalization main study was supported in part by federal funds from the Department of Health and Human Services, Office of the Assistant Secretary for Preparedness and Response; Biomedical Advanced Research and Development Authority, Division of Research Innovation and Ventures under contract 75A50120C00190.

Authors’ Contributions
JP was involved in finalizing the study protocol, implementing the study processes, conducting interviews with participants, analyzing the qualitative data, and drafting the manuscript. JLT was the RiskSEARCH study coordinator responsible for implementing the study processes and drafting the quantitative metrics sections. NZ was involved in implementing the study processes, developing the graphic display for symptoms in the quantitative section, and reviewing the manuscript. AW and MW were involved in finalizing the study protocol and reviewing the manuscript.

Conflicts of Interest
JP, JLT, MW, AW, and NZ are employees of Current Health, a Best Buy company, and receive compensation that supports the conduct of this research.

Multimedia Appendix 1
Inclusion and exclusion criteria to the Risk Stratification and Early Alerting Regarding COVID-19 Hospitalization main study.
Multimedia Appendix 2
Risk Stratification and Early Alerting Regarding COVID-19 Hospitalization interview schedule.

References


Abbreviations

ICF: informed consent form
QSG: QuickStart Guide
RiskSEARCH: Risk Stratification and Early Alerting Regarding COVID-19 Hospitalization
TDF: Theoretical Domains Framework
TUQ: Telehealth Usability Questionnaire
VCT: virtual clinical trial

Edited by A Mavragani; submitted 08.03.22; peer-reviewed by C Carlin, H Karim, V Block, R Ciorap; comments to author 03.05.22; revised version received 19.05.22; accepted 19.05.22; published 05.07.22.

Please cite as:
Pugmire J, Lever Taylor J, Wilkes M, Wolfberg A, Zahradka N
Participant Experiences of a COVID-19 Virtual Clinical Study Using the Current Health Remote Monitoring Platform: Case Study and Qualitative Analysis
JMIR Form Res 2022;6(7):e37567
URL: https://formative.jmir.org/2022/7/e37567
doi:10.2196/37567
PMID:35671408

©Juliana Pugmire, Jessie Lever Taylor, Matt Wilkes, Adam Wolfberg, Nicole Zahradka. Originally published in JMIR Formative Research (https://formative.jmir.org), 05.07.2022. 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.
Measures of Daily Activities Associated With Mental Health (Things You Do Questionnaire): Development of a Preliminary Psychometric Study and Replication Study

Nickolai Titov¹, PhD; Blake F Dear¹, PhD; Madelyne A Bisby¹, PhD; Olav Nielssen¹, PhD; Lauren G Staples¹, PhD; Rony Kayrouz¹, PhD; Shane Cross¹, PhD; Eyal Karin¹, PhD

eCentreClinic, School of Psychological Sciences, Macquarie University, Sydney, Australia

Corresponding Author:
Nickolai Titov, PhD
eCentreClinic
School of Psychological Sciences
Macquarie University
73 Talavera Road
Sydney, 2109
Australia
Phone: 61 0488 991 122
Email: nick.titov@mq.edu.au

Abstract

Background: A large body of research has identified modifiable cognitions and behaviors (actions) associated with psychological health. However, little is known regarding the actions that are most strongly associated with psychological health or the frequency with which they should be performed.

Objective: This paper described 2 studies that used survey methodology to create the Things You Do Questionnaire (TYDQ), which aims to identify and rank actions (items) and domains of actions (factors) most strongly associated with psychological health.

Methods: We used digital marketing strategies to recruit Australian adult participants, who were asked to complete 2 web-based surveys comprising versions of the TYDQ; validated measures of depression, anxiety, and satisfaction with life; and demographic questions. In study 1, a total of 3040 participants rated how often they performed each of the 96 items comprising the TYDQ. This design was replicated in study 2, in which a 59-item version of the TYDQ was completed by 3160 participants. In both studies, the factor structure and validity were examined, as were the associations between individual TYDQ items and 3 mental health outcomes: depression, anxiety, and satisfaction with life.

Results: In study 1, factor analyses revealed that a 5-factor model comprising 27 items achieved an optimum balance between brevity and variance and accounted for 38.1%, 31.4%, and 33.2% of the variance in scores on measures of depression, anxiety, and satisfaction with life, respectively. The factors were interpreted as realistic thinking, meaningful activities, goals and plans, healthy habits, and social connections. These 5 factors were more strongly associated with psychological health than those such as practicing kindness, exercising gratitude, and practicing spirituality. This pattern of results was replicated across gender, age groups, and depression severity. The 5-factor solution found in study 1 was replicated in study 2. Analyses revealed that a 21-item version accounted for 46.8%, 38.2%, and 38.1% of the variance in scores on measures of depression, anxiety, and satisfaction with life, respectively.

Conclusions: These findings indicate that some actions are more strongly associated with psychological health than others and that these activities fall within 5 broad domains, which represent skills often taught in psychological treatments. Subsequent studies are planned to explore the reliability of these items and results in other samples and to examine patterns of change in scores during treatment for anxiety and depression. If replicated, these efforts will assist in the development of new psychological interventions and provide an evidence base for public mental health campaigns designed to promote good mental health and prevent the emergence of common mental disorders.

(JMIR Form Res 2022;6(7):e38837) doi:10.2196/38837
KEYWORDS

anxiety; depression; satisfaction with life; COVID-19; behavior; habits; cognitions; survey; mechanisms; psychological well-being

Introduction

Depressive and anxiety disorders are highly prevalent conditions associated with significant distress, disability, and reduced quality of life [1,2]. These conditions exist along a continuum of severity, and it is now recognized that subclinical or subsyndromal variants can also cause functional disability [3] and considerable economic costs [4]. These disorders are characterized by maladaptive patterns of cognition, behavior, habits, and interpersonal and social engagement, which form the basis for accepted diagnostic criteria [5,6]. Maladaptive patterns of thinking and behavior trigger symptoms and impede recovery. Although the specific mechanisms of change during psychotherapy are not well understood [7-9], the premise that these maladaptive patterns can be modified or changed underlies the teaching of psychological skills within several treatment approaches, including cognitive therapy [10], behavior therapy [11,12], mindfulness [13], and interpersonal therapy [14].

Therefore, identifying adaptive and maladaptive patterns is a priority during treatment, and several approaches have been used to deconstruct these patterns into measurable and modifiable cognitions and behaviors (actions). One approach involves creating questionnaires to measure the quality and frequency of use of cognitive and behavioral skills (cognitive behavioral therapy) taught in psychological treatments [15-21]. Examples of items in these questionnaires include I made an effort to evaluate my negative thoughts by considering just the facts [19] and I worked on a project that was meaningful to me [21]. Consistent with psychological models, these and other studies have found small to moderate relationships between the frequency and quality of cognitive behavioral therapy–related actions and symptom severity over treatment. Although useful in clinical settings, disadvantages with many of these questionnaires include their tendency to focus solely on skills taught in therapy, their frequent use of highly technical language to describe cognitions and behaviors, and the exclusion of a large range of other activities that might be linked to mental health outcomes and that might represent important targets for intervention.

Other researchers surveyed consumers and mental health professionals to generate lists of actions believed to affect the symptoms of anxiety and depression [22-26]. Examples of actions identified in these studies include focusing on the positive [25] and doing more things you enjoy [22]. Many of the actions identified in this research are consistent with the skills taught and promoted in psychological treatments. Unsurprisingly, studies have also shown that prompting these types of actions is associated with symptom reduction [27]. However, a weakness of many of these measures is that they fail to inquire about the frequency of performing such actions. Such information can be important for developing interventions to improve mental health or prevent mental ill health.

An additional and parallel body of work has sought to examine the relationship between mental health and quality of life. This has led to the development of numerous measures, such as the Quality of Life Inventory [28] and the World Health Organization Quality of Life Scale [29], which have been demonstrated to reliably and validly measure satisfaction across multiple life domains (eg, physical health, psychological health, and relationships) while being reliable across both general and clinical populations. More recent quality of life measures, such as the Brunnsviken Brief Quality of Life Scale [30], have aimed to create freely accessible and brief versions that maintain strong validity and reliability profiles. Although helpful in determining satisfaction levels across different life domains, quality of life measures have not been designed to inquire about the frequency with which people engage in modifiable activities, which limits the utility of these tools.

Several research domains have identified actions that are associated with psychological health; however, several important questions remain unanswered. First, we do not know how often actions need to be performed each day or week for a person to experience an improvement in psychological health. Second, we do not know which actions or domains of actions are most strongly associated with improved mental health. Third, research has not yet explored whether particular actions are more helpful for different people according to their clinical presentation, personality, or demographic characteristics.

The studies presented in this paper are part of a research program, the Things You Do Project, designed to examine these questions with the overall aim of identifying and confirming the everyday actions that are most strongly associated with good mental health. The potential value of this information is not only in the treatment of mental disorder but also in the evidence-based prevention of mental disorder and promotion of mental health at the population level. Consistent with the World Health Organization’s definition of health as not only the absence of symptoms but also the presence of a positive state of well-being [31], we used the term psychological health in this paper to describe not only the absence or reduction in symptoms of anxiety or depression but also an improvement in well-being, measured based on self-reported satisfaction with life.

In this paper, we have described 2 studies that used survey methodology to create the Things You Do Questionnaire (TYDQ), which aims to identify and rank actions and domains of actions most strongly associated with psychological health. Recognizing the complexity of this objective and the range of potential methods, we opted for a broad and systematic approach to scale development [32].

In the first study, participants rated the frequency with which they completed a range of actions previously identified to be associated with psychological health. Responses were analyzed to assess the factor structure, reliability, validity, and strength of association of TYDQ items with psychological health. In the second study, participants rated a shorter version of the TYDQ. In both studies, participants also completed validated measures of depression, anxiety, and satisfaction with life. Formal
hypotheses were not proposed, but we expected that the process of ranking actions and domains of actions would generate a parsimonious list of psychological health items.

**Study 1**

The aims of study 1 were to (1) generate a large list of actions previously associated with psychological health and explore the relationship between the weekly frequency of performing these actions and symptoms of anxiety, depression, and life satisfaction; (2) explore the underlying factor structures of the TYDQ and rank the association between the actions and factors and psychological health; (3) explore the importance of different actions across different subgroups; and (4) develop a parsimonious list of items.

**Methods**

**Ethics Approval**

A web-based survey design was used for data collection. The study was reviewed and approved by the Macquarie University Human Research Ethics Committee (MQ HREC: 5201700988), and informed consent was obtained from all participants. All methods were performed in accordance with the relevant guidelines and regulations.

The data sets used and analyzed during this study are available from the corresponding author (NT) upon reasonable request.

**Questionnaire and Item Generation**

An initial list of items was derived from a review of the literature on actions associated with psychological skills identified through multiple sources. These included reviews of (1) skills typically taught in psychological therapies (eg, cognitive skills, behavioral activation, exposure, goal setting, problem solving, social skills training, applied relaxation, mindfulness, acceptance, gratitude, and kindness); (2) actions associated with psychological health, such as healthy daily routines (eg, sleep hygiene, nutrition, and moderated use of electronic devices), physical health (eg, exercise, yoga, stretching, and walking), social activity (eg, talking to friends and spending time with people), meaningful and satisfying activities (eg, doing fun activities), and spiritual and religious activities; (3) symptoms of high prevalence mental disorders as described in psychiatric diagnostic systems [5,6]; (4) previous questionnaires developed by the authors [20], other questionnaires, or similar lists [17-19,22,23]; and (5) consultations with colleagues.

Our research team, consisting of psychiatrists, clinical psychologists, and data analysts with experience in questionnaire development, created a preliminary list comprising >500 actions. Each action was phrased as a verb, and references to psychological techniques or orientations, such as practicing mindfulness, were avoided. During the planning phase, we met regularly to identify, review, edit, and categorize the list, which was reduced to 272 items after duplicates and overlapping items were removed. The items were categorized according to the mechanisms or processes associated with each item; that is, a priori attempts were made to identify theoretical clusters. For example, some items were categorized as primarily concerned with cognitive actions (primary cluster), and within this cluster, they were subdivided into actions associated with challenging unhelpful thoughts, focusing on the future (secondary clusters). In instances where items belonged to multiple clusters, categorization was based on consensus. To capture a broad range of items, the list was shared with colleagues outside the research group, who made further recommendations. The final list of 96 items was based on the following criteria: (1) actions that can be performed daily; (2) measurable, that is, a person can be expected to reliably estimate the frequency of actions over a week; (3) can be completed by most adults; (4) described as a positive action, that is, an action that can be considered a strength; (5) modifiable, that is, can be changed by the person; and (6) did not duplicate another item.

A 5-point Likert rating scale was used to ask participants how often they performed these behaviors over the past week by using the following scoring system: 0=Not at all (0 days per week); 1=1 or 2 days per week; 2=3 to 4 days per week; 3=5 to 6 days per week; and 4=every day (7 days per week). None of the items were reverse scored. After completing the TYDQ, participants were also invited to nominate other actions that were not included in the 96 items they used to improve their mental health.

**Participants and Procedure**

The survey was promoted to adults across Australia via websites and web-based newsletters from Australian mental health services and Facebook posts and advertisements. Advertisements invited people to participate in a study that aimed to identify activities and habits that affect mental health. Consenting participants provided demographic details and completed web-based questionnaires. Participants were required to be aged ≥18 years. No other exclusion criteria were applied. Details of the sample are included in the Results section.

**Measures**

Three standardized outcome measures were used to evaluate and rank the association of the items with psychological well-being.

**Patient Health Questionnaire 9-Item Scale**

The Patient Health Questionnaire 9-item (PHQ-9) scale measures the occurrence of Diagnostic and Statistical Manual IV–congruent depressive symptoms over the past 2 weeks by using a 4-point Likert scale ranging from 0 (not at all) to 3 (nearly every day). Higher scores indicate greater symptom severity [33]. The PHQ-9 has good internal consistency and is sensitive to change [34], with a total score of ≥10 usually but not always associated with a diagnosis of major depressive disorder [35,36]. The Cronbach α in this study was .90.

**Generalized Anxiety Disorder 7-Item Scale**

The Generalized Anxiety Disorder 7-item (GAD-7) scale measures the occurrence of general anxiety symptoms over the past 2 weeks by using a 4-point Likert scale ranging from 0 (not at all) to 3 (nearly every day). The GAD-7 is sensitive to Diagnostic and Statistical Manual IV–congruent GAD, social phobia, and panic disorder, with increasing scores indicating greater severity of symptoms [34,37]. The GAD-7 has been shown to demonstrate sound psychometric properties, with a
total score ≥10 usually indicating a likely diagnosis of an anxiety disorder [38]. The Cronbach α in this study was .92.

**Satisfaction With Life Scale**

The Satisfaction With Life Scale (SWLS) is a 5-item scale that measures attitudes toward life satisfaction by using a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree), with higher scores indicating greater life satisfaction. The SWLS has demonstrated good psychometric properties [39,40] and has been used extensively as a measure of life satisfaction in mental health research [41]. The Cronbach α of the SWLS in this study was .90.

**Statistical Analyses**

The examination of the relationship between the TYDQ items and the standardized scales was operationalized in three detailed and comprehensive steps: (1) item-level analysis, (2) factor analysis, and (3) generalizability of factor solutions.

**Item-Level Analysis**

First, a series of univariate regression models (ANOVA) was used to quantify and statistically test the association between each of the TYDQ item scores and the PHQ-9, GAD-7, and SWLS outcome scores. The strength of the association between items and outcomes was summarized with an $R^2$ metric, noting the proportion of the variance explained by each TYDQ item and each outcome. The $R^2$ values were used to weigh and rank the 96 items.

To further visualize the relative importance of different TYDQ items, each possible TYDQ item score, ranging from 0 (not at all) to 4 (daily), was tabulated with its corresponding PHQ-9, GAD-7, and SWLS outcomes to form a heat map [42]. Comparing the corresponding outcome scores with the range of TYDQ values enabled the identification of the minimal weekly behavioral frequency score needed to achieve optimal PHQ-9, GAD-7, and SWLS outcome scores for each TYDQ item. The optimal TYDQ score was defined as the minimum weekly behavioral frequency beyond which no additional statistically significant improvement can be observed.

**Factor Formation and Analysis**

Second, a series of exploratory factor analyses (EFAs) was used to examine how different TYDQ items formed composite latent factors of behavior. An initial EFA was used to identify the factors based on a full list of 96 items. An additional 3 EFAs were conducted to identify latent factors among the subsets of TYDQ items that were associated more closely with the PHQ-9, GAD-7, and SWLS. These EFAs analyzed the subsets of items based on a minimum of 5%, 10%, and 15% $R^2$ relationship to any one mental health outcome with the aim of identifying more parsimonious lists of factors closely related to psychological health. A factor was considered viable when the item set comprised a composite eigenvalue greater than 1. Additional diagnostic analyses were used to assess each of the identified factors, including item reliability analysis, mean item intercorrelation, and the ability of the identified factors to replicate each item (ie, commonalities).

All identified factors were tested and ranked for their strength of association with the 3 outcomes by using the same analytics and visualization approach used to evaluate TYDQ items. To further evaluate the importance of different behavioral TYDQ factors for mental health, a series of binary logistic regressions was conducted to test the association between the factor scores and the probability of an individual presenting with PHQ-9 or GAD-7 scores that would be considered within the clinical range.

Each factor was assigned a proposed label, which aimed to represent the overall theme of those items. Consensus regarding the factor label was achieved through discussion and debate, considering other labels could have been used.

**Generalizability of Factors Solutions Across Different Subgroups**

A series of confirmatory factor analyses (CFAs) were conducted to examine the reliability of the EFAs within a sample cross-validation analysis (5 randomized subsamples) and the validity of the factor structure along with key subgroups that differed based on age (<30, 30-44, 45-60, >60 years), gender (male, female, and other), the severity of depressive or anxiety symptoms (minimal-mild, moderate, and severe), an indication of employment, and tertiary education. These CFA tests followed the methodology and reporting conventions outlined by Putnick and Bornstein [43] for the examination of measurement invariance, including the examination of configural invariance, metric invariance, scalar invariance, and strict invariance. From the viewpoint of scale development, this step sought to evaluate the reliability and generalizability of the identified factor solutions along relevant subgroups in clinical research (dimensionality).

All analyses were performed using R statistical software version 4.1.1 [44] and the Lavaan package [45]. A conservative $P$ value of .01 was considered the threshold for statistical significance, reflecting the large sample and the multiple number of tests conducted. In all analyses, participants with missing data were not included in the analyses owing to difficulty imputing unbiased missing outcomes in a high-dimensional analysis and cross-sectional settings. All EFAs and CFAs were based on weighted least-squares estimators to account for the ordinal categorical TYDQ item score range. For all EFAs, with the aim of identifying unique subsets of items, a varimax factor rotation was adopted to maximize the number of factors identified and minimize item-factor cross-loadings [46].

Items with factor loadings <0.50, a threshold that aimed to achieve a conservative balance between recommendations in the literature [32,47], were suppressed as were items with suboptimal association with their corresponding factor (accounting for <25% of the factor variance) [48].

**Results**

**Item and Sample Description**

Participants in study 1 were recruited from March to June 2018, during which time 3755 people consented to participate and 3040 (80.96%) completed the questionnaire. Sample characteristics are presented in Table 1. Participants had a broad...
range of both demographics (age, gender, location, education, and employment) and mental health symptoms. Approximately half of the sample scored above the clinical cutoffs on the PHQ-9 (n=1710, 48.1%), GAD-7 (n=2020, 57.9%), and SWLS (n=2188, 64%), allowing the testing of the association of everyday behaviors for those above and below recognized clinical cutoffs on the PHQ-9 and GAD-7.

Multimedia Appendix 1 includes each of the 96 items ordered based on the a priori primary and secondary theoretical clusters created during item development and based on the observed average weekly score. Multimedia Appendix 1 also uses a heatmap to indicate the frequency with which participants reported performing each item each week. For example, the first item in Multimedia Appendix 1 was the 13th item presented in the survey and read I read, listened, or watched something I enjoyed. A priori, this item was categorized as an Activity/meaning (primary and secondary clusters), and the average weekly score was 2.72 (mean TYDQ weekly score), indicating that the mean frequency reported was between 4 and 6 days each week. The final column indicates that 37.4% (n=1404) of the sample reported doing this action daily, whereas 3.7% (n=139) reported not doing this action in the previous 7 days.

As shown in Multimedia Appendix 1, the frequency of actions (TYDQ mean scores) ranged across and within clusters. For example, 2 of the highest frequency scores were observed within the Healthy Routine cluster, with daily showers or baths reported by 69.3% (n=2602) of the sample, and avoidance of illicit drugs or misuse of medication reported by 83.3% (n=3128).
### Table 1. Sample characteristics of studies 1 and 2.

<table>
<thead>
<tr>
<th>Variable and subgroup</th>
<th>Study 1 (year 2018; N=3755)</th>
<th>Study 2 (year 2020; N=3756)</th>
<th>Test of sample differences</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Survey completion, n (%)</strong></td>
<td></td>
<td></td>
<td>( \chi^2_{1,7510}=13.1, P&lt;.001 )</td>
</tr>
<tr>
<td>Completed survey</td>
<td>3040 (81)</td>
<td>3160 (84.1)</td>
<td></td>
</tr>
<tr>
<td>Incomplete survey</td>
<td>715 (19)</td>
<td>596 (15.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td>( \chi^2_{2,7509}=303.8, P&lt;.001 )</td>
</tr>
<tr>
<td>Male</td>
<td>1806 (48.1)</td>
<td>1072 (28.5)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1924 (51.2)</td>
<td>2651 (70.6)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>25 (0.7)</td>
<td>33 (0.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td>( t_{7509}=16.13, P&lt;.001 )</td>
</tr>
<tr>
<td>Value, mean (SD)</td>
<td>41.9 (13.9)</td>
<td>47.7 (17.1)</td>
<td></td>
</tr>
<tr>
<td>&lt;30, n (%)</td>
<td>976 (26)</td>
<td>944 (25.1)</td>
<td></td>
</tr>
<tr>
<td>30-45, n (%)</td>
<td>1416 (37.7)</td>
<td>1430 (38.1)</td>
<td></td>
</tr>
<tr>
<td>45-60, n (%)</td>
<td>1137 (30.3)</td>
<td>925 (24.6)</td>
<td></td>
</tr>
<tr>
<td>&gt;60, n (%)</td>
<td>226 (6)</td>
<td>456 (12.1)</td>
<td></td>
</tr>
<tr>
<td><strong>PHQ-9(^a) category, n (%)</strong></td>
<td></td>
<td></td>
<td>( \chi^2_{1,7095}=159.1, P&lt;.001 )</td>
</tr>
<tr>
<td>Below cutoff (≥10)</td>
<td>1710 (48.1)</td>
<td>2227 (62.9)</td>
<td></td>
</tr>
<tr>
<td>Above cutoff (&lt;10)</td>
<td>1848 (51.9)</td>
<td>1311 (37.1)</td>
<td></td>
</tr>
<tr>
<td><strong>GAD-7(^b) category, n (%)</strong></td>
<td></td>
<td></td>
<td>( \chi^2_{1,6961}=151.3, P&lt;.001 )</td>
</tr>
<tr>
<td>Below cutoff (≥10)</td>
<td>1471 (42.1)</td>
<td>2497 (71.9)</td>
<td></td>
</tr>
<tr>
<td>Above cutoff (&lt;10)</td>
<td>2020 (57.9)</td>
<td>974 (28.1)</td>
<td></td>
</tr>
<tr>
<td><strong>SWLS(^c) category, n (%)</strong></td>
<td></td>
<td></td>
<td>( \chi^2_{1,6858}=98.4, P&lt;.001 )</td>
</tr>
<tr>
<td>Below cutoff (≥16)</td>
<td>2188 (64)</td>
<td>2578 (75)</td>
<td></td>
</tr>
<tr>
<td>Above cutoff (&lt;16)</td>
<td>1233 (36)</td>
<td>860 (25)</td>
<td></td>
</tr>
<tr>
<td>Depression symptoms (PHQ-9), mean (SD)</td>
<td>10.7 (6.8)</td>
<td>8.5 (6.5)</td>
<td>( t_{7095}=14.3, P&lt;.001 )</td>
</tr>
<tr>
<td>Anxiety symptoms (GAD-7), mean (SD)</td>
<td>8.8 (5.8)</td>
<td>6.8 (5.6)</td>
<td>( t_{6961}=14.17, P&lt;.001 )</td>
</tr>
<tr>
<td>Satisfaction with life (SWLS), mean (SD)</td>
<td>18.4 (7.7)</td>
<td>20.8 (7.8)</td>
<td>( t_{6858}=13.42, P&lt;.001 )</td>
</tr>
<tr>
<td><strong>Australian born, n (%)</strong></td>
<td></td>
<td></td>
<td>( \chi^2_{1,6049}=32.9, P&lt;.001 )</td>
</tr>
<tr>
<td>No</td>
<td>632 (16.8)</td>
<td>829 (22.1)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3123 (83.2)</td>
<td>2927 (77.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Location, n (%)</strong></td>
<td></td>
<td></td>
<td>( \chi^2_{2,7392}=4.9, P=.09 )</td>
</tr>
<tr>
<td>Capital city</td>
<td>1921 (51.2)</td>
<td>2015 (54.5)</td>
<td></td>
</tr>
<tr>
<td>Other urban region</td>
<td>934 (24.9)</td>
<td>898 (24.3)</td>
<td></td>
</tr>
<tr>
<td>Rural or remote</td>
<td>841 (22.4)</td>
<td>785 (21.2)</td>
<td></td>
</tr>
<tr>
<td><strong>Education attained, n (%)</strong></td>
<td></td>
<td></td>
<td>( \chi^2_{3,7391}=75.9, P&lt;.001 )</td>
</tr>
<tr>
<td>High school or less</td>
<td>744 (19.8)</td>
<td>589 (15.9)</td>
<td></td>
</tr>
<tr>
<td>Other tertiary qualification or certificate</td>
<td>1200 (32)</td>
<td>1100 (29.7)</td>
<td></td>
</tr>
<tr>
<td>Postgraduate degree</td>
<td>735 (19.6)</td>
<td>1040 (28.1)</td>
<td></td>
</tr>
<tr>
<td>Undergraduate degree</td>
<td>1017 (27.1)</td>
<td>969 (26.2)</td>
<td></td>
</tr>
<tr>
<td><strong>Employment status, n (%)</strong></td>
<td></td>
<td></td>
<td>( \chi^2_{5,7389}=245.9, P&lt;.001 )</td>
</tr>
</tbody>
</table>

---

\(a\) PHQ-9 = Patient Health Questionnaire-9

\(b\) GAD-7 = Generalised Anxiety Disorder-7

\(c\) SWLS = Satisfaction with Life Scale
<table>
<thead>
<tr>
<th>Variable and subgroup</th>
<th>Study 1 (year 2018; N=3755)</th>
<th>Study 2 (year 2020; N=3756)</th>
<th>Test of sample differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paid employment</td>
<td>2335 (62.2)</td>
<td>2031 (54.9)</td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>383 (10.2)</td>
<td>285 (7.7)</td>
<td></td>
</tr>
<tr>
<td>Home duties or parenting</td>
<td>177 (4.7)</td>
<td>154 (4.2)</td>
<td></td>
</tr>
<tr>
<td>Disability support</td>
<td>196 (5.2)</td>
<td>188 (5.1)</td>
<td></td>
</tr>
<tr>
<td>Unemployed or seeking employment</td>
<td>318 (8.5)</td>
<td>289 (7.8)</td>
<td></td>
</tr>
<tr>
<td>Retired</td>
<td>287 (7.6)</td>
<td>751 (20.3)</td>
<td>$\chi^2_{3,7391}=119.4, P&lt;.001$</td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single or ever married</td>
<td>996 (26.5)</td>
<td>1125 (30.4)</td>
<td>$\chi^2_{3,7394}=60.0, P&lt;.001$</td>
</tr>
<tr>
<td>Widowed</td>
<td>55 (1.5)</td>
<td>145 (3.9)</td>
<td></td>
</tr>
<tr>
<td>Separated or divorced</td>
<td>378 (10.1)</td>
<td>555 (15)</td>
<td></td>
</tr>
<tr>
<td>Married or defacto</td>
<td>2267 (60.4)</td>
<td>1873 (50.6)</td>
<td></td>
</tr>
<tr>
<td>Seeing mental health professional about anxiety or depression? n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>747 (19.9)</td>
<td>999 (27)</td>
<td>$\chi^2_{1,7451}=18.5, P&lt;.001$</td>
</tr>
<tr>
<td>Previously</td>
<td>1737 (46.3)</td>
<td>1714 (46.3)</td>
<td></td>
</tr>
<tr>
<td>Currently</td>
<td>1212 (32.3)</td>
<td>985 (26.6)</td>
<td></td>
</tr>
<tr>
<td>Taking medication for anxiety or depression? n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2356 (62.7)</td>
<td>2496 (67.5)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1399 (37.3)</td>
<td>1202 (32.5)</td>
<td></td>
</tr>
</tbody>
</table>

aPHQ-9: Patient Health Questionnaire 9-item.  
bGAD-7: Generalized Anxiety Disorder 7 item.  
cSWLS: Satisfaction With Life Scale.

**Item-Level Analysis**

The 96 items were then subjected to tests of the strength of association with each of the PHQ-9, GAD-7, and SWLS outcomes. The results from this first series of analyses are presented in Multimedia Appendix 2 for each item, with items grouped around the a priori clusters. The columns of the table report the mean PHQ-9, GAD-7, and SWLS scores observed with differing rates of weekly behavioral frequency. The PHQ-9, GAD-7, and SWLS scores were further incorporated into a heat map visualization to highlight the TYDQ items associated with the highest symptom scores among the 96 TYDQ items, where red hues indicate deterioration of psychological health (ie, symptom scores and satisfaction with life) and blue hues indicate improved psychological health.

The proportion of symptom variance explained ($R^2$) was reported as a summary measure of the association between each item and outcome. In Multimedia Appendix 2, the presentation order of the 96 items illustrates the relative ability of TYDQ items to account for the cumulative variance of the 3 outcomes, with the strongest items reported at the top of each a priori cluster. Together, the 96 items were significantly associated with each of the 3 outcomes (TYDQ outcome regression $\beta$ test resulting in $P<.001$), except item TYDQ 57 (*I fulfilled my responsibilities even though I didn’t want to*) and its association with GAD-7 outcomes. Most items showed a linear relationship between higher frequency and lower symptom scores, with a minority of items showing a flattened, convexed, or positive relationship with PHQ-9 and GAD-7 outcomes. However, the strength of the association between each of the items and each outcome varied substantially, both within and across clusters. For example, TYDQ 16, *I treated myself with respect*, illustrated the strongest association with depressive symptoms (accounting for 22.5% of all PHQ-9 scores); TYDQ 70, *I kept a realistic perspective on things*, was the strongest correlate of anxiety symptoms (accounting for 21.1% of all GAD-7 scores); and TYDQ 54, *I had something to look forward to*, was the strongest correlate of satisfaction with life (accounting for 25.9% of all SWLS scores). Across the items, the threshold of activity associated with optimal well-being varied from half a week to daily, with the majority of items requiring frequent (almost every day) but not necessarily daily activity.

**Factor Formation and Analysis**

EFAs were conducted to identify underlying groupings among the items and to explore a brief, latent factor structure of behaviors associated with psychological health. The results from the EFA of the complete 96-item list are presented in Multimedia Appendix 3. In total, 16 factors were identified, with 56 of the items demonstrating factor loadings above 0.5 for any 1 factor. Factors with loadings of $\leq$0.5 were suppressed. Across the TYDQ list, the factors appeared to form along their a priori clusters, with most items grouped within their clusters. No item-factor cross-loadings were identified above the loading of 0.5.
The results of the tests of association between the resulting factor scores and each of the PHQ-9, GAD-7, and SWLS outcomes are further reported in Multimedia Appendix 3, which describes the outcomes of a series of 4 EFAs, each testing a different number of items (see Multimedia Appendix 3 for items). This shows that the complete list of 16 factors, based on 56 items, accounted for 41%, 37%, and 32% of the outcome $R^2$, respectively. The resulting 16 factors can also be seen as varying in their strength of association with the 3 outcomes, with estimates of $R^2$ ranging from <0.1% to 24%.

The EFA was repeated with another subset of items (68 items), selected based on the individual item’s ability to account for at least 5% of any of the 3 symptom outcomes ($R^2 > 5\%$), which provided an opportunity to examine patterns among TYDQ items more closely related to psychological health. The analysis identified 11 factors, with only 45 items demonstrating factor loadings >0.5 for any one factor. The items were clustered within their respective a priori clusters, and no item-factor cross-loadings were identified. A test of the association between the resulting factor scores and each of the PHQ-9, GAD-7, and SWLS outcomes illustrated that a complete list of 11 factors based on 45 items accounted for 33%, 29%, and 32% of the outcome $R^2$, respectively, with each of the individual factors varying in their strength of association from 5% to 21% (Multimedia Appendix 4).

In a third EFA, an even briefer subset of items (35 items) was selected based on the item’s ability to account for a minimum of 10% of any of the 3 symptom outcomes. This resulted in 5 factors and 27 items with a factor loading above 0.5 for any 1 factor. No item-factor cross-loadings were identified. A test of association between the resulting factor scores and each of the PHQ-9, GAD-7, and SWLS outcomes revealed a list of 5 factors based on 27 items, which accounted for 38%, 31%, and 33% of the outcomes $R^2$, respectively (Multimedia Appendix 4). The formation of factors largely followed the a priori clusters, and the resulting factors varied 3-fold in their strength of association with each of the outcomes (9%-28%).

A fourth EFA, which included items, identified a stringent $R^2 > 15\%$ criterion. This resulted in a brief set of items (8 items) found to represent a single composite factor combining key items from several domains. Although this result is interesting, the PHQ-9, GAD-7, and SWLS variance explained by a single factor (Multimedia Appendix 4) was suboptimal ($R^2 = 36.6\%$, $R^2 = 26.7\%$, and $R^2 = 32.9\%$, respectively), and the representation of multiple domains, including activity planning, self-representation, and values through a single nonspecific factor, would inevitably limit the ability to assess and interpret the type of behavior.

Together, these EFAs illustrate that the 96-item list could be reduced to a more parsimonious and selective item list, with an optimal 5-factor, 27-item solution that retained >90% of the strength of association with each of the symptom outcomes, with only a quarter of the items (27/96, 28%). Each factor solution was further assessed for item reliability, mean item intercorrelation within each factor, and the ability of the identified factors to replicate each item (commonalities), as collated in Multimedia Appendix 5.

**Generalizability of Factor Solutions Across Different Subgroups**

A series of CFAs were conducted to examine the measurement invariance of the 5-factor solution as a brief but sensitive measure of psychological well-being across 5 categories, namely age groups, gender, depression and anxiety symptoms, education, and employment. The results from the CFAs are collated in Multimedia Appendix 6. The identification of TYDQ differences in the intercept and residual scores indicated that gender, age, and baseline symptoms were associated with different frequencies of weekly activity, although the TYDQ scores formed the same latent patterns across these groups. To check for the potential of alternate models, we reran the EFA within each of the age, sex, and PHQ-9 subgroups. The results identified the 5-factor solution reliably replicated as the most prominent latent solution when no parameter restraints were made or when the rotation methodology was changed (Multimedia Appendix 7). Together, the results from these CFAs illustrate that the latent 5-factor structure identified was statistically reliable, generalized across subgroups of clinical interest, and replicated across differing aspects of the statistical methodology (rotation and parameter constraint).

**Discussion**

Study 1 involved a comprehensive series of methodological and statistical steps designed to create a large list of actions associated with psychological health, test their strength of association with defined mental health outcomes, and develop a parsimonious list of items. To begin with, 96 items previously associated with psychological health were identified and administered to a large and diverse adult sample, along with measures of depression, anxiety, and satisfaction with life.

Several factor solutions were explored, and through a systematic process of dimension reduction, EFA solutions based on 56 items, 45 items, 27 items, and 8 items were examined. The 27-item solution, which comprised 5 factors, was found to be optimal, as it represented 38%, 31%, and 33% of the variance in scores of measures of depression, anxiety, and satisfaction with life, respectively. A series of CFAs indicated that the factor structure was robust across different subgroups, with these 5 factors interpreted as (1) realistic thinking, (2) meaningful activities, (3) goals and plans, (4) healthy habits, and (5) social connections.

Surprisingly, some items and factors that were expected to show strong relationships with psychological health, such as actions associated with kindness, gratitude, and spirituality, were not represented in the 5-factor solution. In contrast, items associated with skills commonly taught in psychological treatments, such as challenging unhelpful thoughts and engaging in pleasant activities, were very strongly associated with psychological health. Importantly, the results of study 1 not only identified items and factors that are strongly associated with psychological health but also found that the frequency of performing those actions was associated with improved well-being.
In summary, the results of study 1 provided preliminary psychometric evidence for the TYDQ as a measure of modifiable actions associated with psychological health. The reliability of the strong relationship between the 5-factor 27-item list found in study 1 needs to be replicated in further studies, which is one of the aims of study 2. Given the surprising finding that some items and factors were not strongly associated with psychological health, such items should also be retested.

**Study 2**

**Overview**
The objective of study 2 was to explore the reliability of the results of study 1 by replicating the design used in study 1 using a briefer version of the TYDQ. The aims were to (1) generate a shortened list of items based on the results of study 1 and explore the relationship between the weekly frequency of performing these items and outcomes of psychological health; (2) explore the underlying factor structures and rank the association between the items, factors, and psychological health; (3) explore the importance of different actions across different subgroups; and (4) develop a parsimonious list of items.

It should be noted that study 2 was conducted when major cities across Australia were locked down in response to the COVID-19 pandemic, which, on the one hand, represents a significantly changed context to that of study 1 but, on the other hand, provides an opportunity to test the items and the questionnaire in a psychologically challenging context.

**Methods**

**Ethics Approval**
The design of study 2 replicated that used in study 1. Ethics approval for data collection was obtained from the Macquarie University Human Research Ethics Committee (MQ HREC: 5201700988), and informed consent was obtained from all participants.

**Questionnaire**
Study 2 evaluated a version of the TYDQ comprising 59 of the 96 original items (Multimedia Appendix 8), with each item using the same 5-point rating scale used in study 1. This brief list was constructed in several steps. First, items that accounted for at least 5% of the PHQ-9, GAD-7, and SWLS were identified, resulting in an initial list of 67 items. Second, this list was edited to remove items with similar wording or duplicated items, resulting in the removal of 23 items. Third, 15 additional items from the original 96 item list were added to test the association between items, such as kindness to others and gratitude, which were considered clinically important and expected in study 1 to be associated with the outcomes but were not. We reviewed each of these steps.

**Participants and Procedure**
The procedure used in study 1 was repeated in study 2. As shown in Table 1, a total of 3756 participants consented to participate in the study and started filling the questionnaires, which were available between June and August 2020. These
then compared with the same 21 items from the study 2 results, and associations between the 5 factors and the 3 outcome variables were compared across samples (Multimedia Appendix 10). This demonstrated close similarities in strength, directionality, significance, and optimal frequency cutoffs between the 2 samples.

A set of 3 EFAs were then conducted for item subsets with no selection criteria, inclusion criteria of $R^2 > 10\%$, and inclusion criteria of $R^2 > 15\%$. The 5-factor 21-item solution associated with the $R^2 > 10\%$ criteria in study 2 accounted for the highest amount of PHQ-9, GAD-7, and SWLS outcome variance ($R^2 = 46.8\%$, $R^2 = 38.2\%$, and $R^2 = 38.1\%$, respectively). Consistent with this, a reanalysis of the data from study 1 using the same $R^2 > 10\%$ criteria demonstrated that this 5-factor 21-item solution accounted for the highest amount of PHQ-9, GAD-7, and SWLS outcome variance ($R^2 = 37.1\%$, $R^2 = 30.0\%$, and $R^2 = 30.5\%$, respectively).

**Generalizability of Factor Solutions Across Different Subgroups**

CFAs seeking to verify a 5-factor 21-item solution across the study 1 and study 2 samples resulted in metric invariance (similarities in factors and item loading) but not scalar or strict invariance (Multimedia Appendix 11). This result is consistent with the interpretation that participants from the 2 studies differed in their means but not item or factor importance.

In brief, the overall results of study 2 replicated the results and 5-factor model obtained in study 1, albeit with a 21- rather than a 27-item solution. These results confirm the strength of the relationship between key items in the TYDQ and psychological health outcomes and the underlying factor structure.

**Discussion**

Study 2 examined the performance of a shortened version of the TYDQ. The survey was completed by 3160 people, and although the sample was significantly different from that of study 1, the overall patterns of results obtained in study 1 were replicated. The 5-factor structure observed in study 1 was supported and accounted for 46.8\%, 38.2\%, and 38.1\% of the variance in the symptoms of depression, anxiety, and satisfaction with life, respectively. These solutions were achieved using a parsimonious 21-item list. Importantly, and again consistent with the results of study 1, the observations that some items and factors expected to show strong relationships with psychological health, such as showing kindness to others and expressing gratitude, were not found to be associated with well-being. The implications and limitations of these results are discussed in the subsequent section.

**General Discussion**

This paper describes 2 studies reporting on the initial development and evaluation of the TYDQ, which aims to measure modifiable actions that are strongly associated with psychological health. The detailed and systematic approach followed accepted scale development methods [32] and attempted to address the limitations of previous work by including a broad range of modifiable actions, exploring their frequency over a 1-week timeframe, identifying the actions most strongly related to the target outcomes, and testing the generalizability of results across age, gender, and baseline symptom severity.

**Principal Findings**

Study 1 examined the performance of a 96-item version of the TYDQ. A 27-item 5-factor solution achieved an optimal balance between accounting for sufficient variance in outcome measures of depression, anxiety, and satisfaction with life and parsimony. These five factors were interpreted as actions concerned with (1) realistic thinking, (2) meaningful activities, (3) goals and plans, (4) healthy habits, and (5) social connections. There was a strong relationship between the frequency of performing the actions associated with these factors each week and improved psychological health, with results indicating that they should be performed at least half a week or more to optimize their psychological health. Importantly, the factor structure was consistent across demographic variables, including gender and age groups.

Overall, these findings were replicated in study 2, which evaluated a shortened 59-item version by using the same systematic analytic strategy. Study 2 was conducted during a period when participants were mostly in lockdown owing to the impact of the COVID-19 pandemic in Australia. However, the strength of the relationship between the frequency of target actions and outcomes increased relative to study 1. The same 5-factor solution found in study 1 was also found in study 2, although slight differences in the item loadings were observed across the 2 samples. An item-reduction exercise revealed that a 5-factor 21-item version solution achieved an optimal balance between item parsimony and accounting for sufficient variance across the 3 outcome measures. These results appear to confirm the robustness of the 5 factors but also suggest that the 21-item list (Multimedia Appendix 9) should not be considered final, and additional survey design efforts could further improve the brevity and validity of the items identified. In any case, the 96-item and 59-item lists are provided in Multimedia Appendix 11.

**Comparison With Prior Work**

The overall results are consistent with previous research regarding key items associated with psychological health, including cognitions and actions such as, *I did something enjoyable, I kept a realistic perspective on things, I had something to look forward to, and I treated myself with respect.* These 5 factors confirm the importance of many of the skills taught in psychological interventions, including cognitive therapy [10], behavior therapy [12,49], and others. Various combinations of these skills and actions have been recognized as contributing to psychological health [50-52] and are frequently included in public mental health campaigns. Thus, the contribution of our preliminary study is not to identify the actions and factors associated with psychological health. Instead, these studies extend the literature by comparing the relative benefits of different types of actions on psychological health and the frequency with which these factors are associated with psychological health. The findings indicate that performing...
these actions for at least half the days of the week is important for psychological health across age, gender and demographic groups, as well as in different severity levels of mood symptoms. This proposed frequency needs to be carefully evaluated using a longitudinal study; however, if supported, it would indicate a key target for psychological interventions as well as for public health interventions designed to promote psychological health and prevent common mood disorders in the community.

An additional and important finding was that some actions were less strongly associated with psychological health than expected. For example, acts of kindness, practicing gratitude, and spirituality are widely considered important for psychological health; however, in this study, these actions were not found to be as important as those relevant to the 5 factors. This lack of findings is broadly consistent with the results of empirical studies [53-59].

For example, in a systematic review and meta-analysis of 27 experimental studies examining the relationships between acts of kindness and different measures of subjective well-being, including measures of psychological health, Curry et al [53] found the overall effect of kindness to be small to medium (δ=0.28); however, they noted that most of the reviewed studies were underpowered to detect small differences. In a series of meta-analyses examining the findings of 38 studies, Dickens [54] showed that when compared with waitlist, inactive, or no-treatment control groups, gratitude interventions were associated with decreased depressive symptoms and improved positive affect, well-being, happiness, and life satisfaction, albeit with small effect sizes (Cohen d=0.13−0.30). In a more recent review of the effect of gratitude traits and interventions, Jans-Beken et al [60] concluded that gratitude interventions were not found to reliably improve symptoms, such as depression and anxiety, but they were associated with improved happiness and life satisfaction. In a meta-analysis of 48 longitudinal studies, Garssen et al [56] reported an overall positive effect of religion or spirituality on mental health; however, similar to studies examining the relationships between mental health and acts of kindness and practicing gratitude, the effects were small, with a random weighted average effect size of r=0.08.

Given the large number and broad cross-section of people completing the 2 surveys, these findings appear to be robust. However, given the preliminary nature of the 2 studies reported here, we propose that the findings should not be interpreted as such actions are not helpful; rather, it is likely that different actions are important at different stages of psychological health and for particular individuals and age groups. For example, it is commonly recognized in clinical practice that people who are severely depressed are more likely to benefit from actions involving increasing reinforcement and pleasure than from practicing gratitude, but once their mood has lifted, actions such as practicing gratitude and showing kindness to others may help maintain psychological health. In addition, it should be noted that the results obtained here are relevant to the outcomes of interest, namely, symptoms of depression, anxiety, and satisfaction with life, and that choosing other outcome measures may result in other factors and items ascending in importance. This raises the possibility of future research exploring the relationship between everyday actions and a much broader range of mental health outcomes, including the general p factor, which extends beyond internalizing disorders to include externalizing and thought or psychotic disorders [61].

Limitations
The preliminary and cross-sectional nature of the 2 studies reported here are associated with several limitations. First, despite the large sample sizes and consistent results across subsamples, the work should be considered preliminary and tested in other samples, including people with different cultural and demographic backgrounds. In addition, the items and factors identified here need to be tested using longitudinal or interventional research designs to gauge their sensitivity to change and reliability over time. The robustness of the 5 factors should also be tested across differing statistical methods for identifying dimensionality, for example, using different statistical criteria for selecting items or using latent profiling.

We also acknowledge that the initial selection of items used to develop the TYDQ, despite the breadth of our inquiries, involved some subjective choices. We have shared lists to assist independent replication and further development (Multimedia Appendix 12). Related to these limitations, we also acknowledge that the terms we generated to describe the factors were based on our clinical and psychometric decisions, and that other research groups may have generated other labels for each of the factors. We also acknowledge that although most of the actions were relatively simple to understand by participants, some actions were more difficult for participants to measure; for example, the item I treated myself with respect. This raises important questions about how people interpreted some of the items, which may in turn affect both the reliability and validity of the results. This is further complicated by the observation that a person’s interpretation of the questions may change over time and with their experience of the action. Finally, we acknowledge that the focus of this study is primarily on the relationship between actions and symptoms and may not be generalizable to other concepts of psychological well-being, happiness, or thriving [62].

Strengths
The strengths of this study include the development of a questionnaire containing items that were not obviously associated with any specific psychological model or approach, evaluated through a systematic and detailed analytic procedure that included multiple tests of generalizability by using 2 large sample sizes. This procedure contributes to the literature by comparing the relative strengths of several groups of modifiable actions and the minimum frequency required for such actions to affect psychological health. An additional strength is that the results were mostly replicated although the second sample was obtained during a challenging period for the community characterized by social restrictions related to the impacts of the COVID-19 pandemic.

Conclusions
In conclusion, these studies provide preliminary evidence for actions and factors strongly associated with psychological health and the frequency with which they should be performed. Future
studies are planned to further explore the patterns of change in brief versions of the TYDQ, including samples of people seeking treatment for mental health conditions. We hope that these efforts will assist in the development of new psychological interventions and provide an evidence base for public mental health campaigns designed to promote good mental health and prevent the emergence of common mental disorders.

Acknowledgments
The studies reported in this paper were partly funded by donations from the Mental Wheels Foundation, which had no role in study design or analysis.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Items, primary and secondary cluster, item wording, and weekly score dispersion by frequency.
[DOCX File, 37 KB - formative_v6i7e38837_app1.docx ]

Multimedia Appendix 2
A heat map of the strength of association of Things You Do Questionnaire item weekly frequency with Patient Health Questionnaire 9-item, Generalized Anxiety Disorder 7-item, and Satisfaction With Life Scale outcomes, and the behavioral frequency thresholds associated with optimal mental health gains.
[DOCX File, 111 KB - formative_v6i7e38837_app2.docx ]

Multimedia Appendix 3
Exploratory factor analysis solution and assigned factor labels for items grouped with no item selection criteria (96 items), and item grouping under R2>5%, 10%, and 15% outcomes association criteria.
[DOCX File, 46 KB - formative_v6i7e38837_app3.docx ]

Multimedia Appendix 4
A heat map of the identified factor association with Patient Health Questionnaire 9-item, Generalized Anxiety Disorder 7-item, and Satisfaction With Life Scale outcomes, and the behavioral frequency thresholds associated with optimal mental health gains.
[DOCX File, 73 KB - formative_v6i7e38837_app4.docx ]

Multimedia Appendix 5
Exploratory factor analysis diagnostic statistics: item commonalities, factor eigenvalues, factor Cronbach α, and item mean intercorrelations.
[DOCX File, 70 KB - formative_v6i7e38837_app5.docx ]

Multimedia Appendix 6
Invariance test solutions examining dimensionality in study 1 sample.
[DOCX File, 56 KB - formative_v6i7e38837_app6.docx ]

Multimedia Appendix 7
Exploratory factor analysis parameter solution replications across key sample dimensions and methods of rotation methods.
[DOCX File, 62 KB - formative_v6i7e38837_app7.docx ]

Multimedia Appendix 8
Means, study 1 and study 2 sample group differences, and association of Things You Do Questionnaire items with each of the Patient Health Questionnaire 9-item, Generalized Anxiety Disorder 7-item, and Satisfaction With Life Scale outcomes.
[DOCX File, 47 KB - formative_v6i7e38837_app8.docx ]

Multimedia Appendix 9
Exploratory factor analysis solution and assigned factor labels for items grouped with no item selection criteria (96 items), and under R2>10% outcomes association criteria.
[DOCX File, 40 KB - formative_v6i7e38837_app9.docx ]
Multimedia Appendix 10
A heat map of the identified factor association with Patient Health Questionnaire 9-item, Generalized Anxiety Disorder 7-item, and Satisfaction With Life Scale outcomes across study 1 and study 2 samples.

[DOCX File, 70 KB - formative_v6i7e38837_app10.docx]

Multimedia Appendix 11
Invariance test solutions examining dimensionality in study 2 sample.

[DOCX File, 55 KB - formative_v6i7e38837_app11.docx]

Multimedia Appendix 12
Things You Do Questionnaire item lists.

[DOCX File, 41 KB - formative_v6i7e38837_app12.docx]

References


33. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]


44. The R project for statistical computing. R Foundation. URL: https://www.r-project.org [accessed 2022-02-01]


URL: https://formative.jmir.org/2022/7/e38837
doi:10.2196/38837
PMID:35788101

©Nickolai Titov, Blake F Dear, Madelyne A Bisby, Olav Nielssen, Lauren G Staples, Rony Kayrouz, Shane Cross, Eyal Karin. Originally published in JMIR Formative Research (https://formative.jmir.org), 05.07.2022. 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.
Standardizing Primary Health Care Referral Data Sets in Nigeria: Practitioners' Survey, Form Reviews, and Profiling of Fast Healthcare Interoperability Resources (FHIR)

Emeka Chukwu¹,², BEng, MSc; Lalit Garg¹, BSc, PhD; Nkiruka Obande-Ogbuinya¹, BSc, MSc, PhD; Vijay Kumar Chattu⁴,⁵,⁶, MD, MPH, MPhil

¹Department of Computer Information Systems, Faculty of Information Communication Technology, University of Malta, Msida, Malta
²Digital Health Interoperability Network, Abuja, Nigeria
³Department of Science Education, Alex-Ekwueme Federal University Ndufu-Alike, Ikwo, Ebonyi, Nigeria
⁴Department of Occupational Science & Occupational Therapy, Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada
⁵Center for Transdisciplinary Research, Saveetha Institute of Medical and Technological Sciences, Saveetha University, Chennai, India
⁶Department of Community Medicine, Faculty of Medicine, Datta Meghe Institute of Medical Sciences, Wardha, India

Corresponding Author:
Emeka Chukwu, BEng, MSc
Department of Computer Information Systems
Faculty of Information Communication Technology
University of Malta
PG room A24, Level 0
Msida, MSD 2080
Malta
Phone: 356 99330888
Email: nnaemeka_ec@hotmail.com

Abstract

Background: Referral linkages are crucial for efficient functioning of primary health care (PHC) systems. Fast Healthcare Interoperability Resource (FHIR) is an open global standard that facilitates structuring of health information for coordinated exchange among stakeholders.

Objective: The objective of this study is to design FHIR profiles and present methodology and the profiled FHIR resource for Maternal and Child Health referral use cases in Ebonyi state, Nigeria—a typical low- and middle-income country (LMIC) setting.

Methods: Practicing doctors, midwives, and nurses were purposefully sampled and surveyed. Different referral forms were reviewed. The union of data sets from surveys and forms was aggregated and mapped to base patient FHIR resource elements, and extensions were created for data sets not in the core FHIR specification. This study also introduced FHIR and its relation to the World Health Organization's (WHO's) International Classification of Diseases.

Results: We found many different data elements from the referral forms and survey responses even in urban settings. The resulting FHIR standard profile is published on GitHub for adaptation or adoption as necessary to aid alignment with WHO recommendations. Understanding data sets used in health care and clinical practice for information sharing is crucial in properly standardizing information sharing, particularly during the management of COVID-19 and other infectious diseases. Development organizations and governments can use this methodology and profile to fast-track FHIR standards adoption for paper and electronic information sharing at PHC systems in LMICs.

Conclusions: We presented our methodology for profiling the referral resource crucial for the standardized exchange of new and expectant moms’ information. Using data from frontline providers and mapping to the FHIR profile helped contextualize the standardized profile.

(JMIR Form Res 2022;6(7):e28510)   doi:10.2196/28510

KEYWORDS

FHIR; COVID-19; digital health; eHealth; mHealth; BlockMom; Nigeria; primary health care; health information; health information exchange; interoperability
Introduction

Background

Health care is a complex sector that involves medical professionals, allied health workers, the information and communication technology (ICT) workforce, and various other stakeholders. The World Health Organization (WHO) highlights the importance of 6 building blocks of any health system: service delivery, health workforce, health information systems, access to essential medicines, financing, and leadership or governance [1]. Therefore, the health information systems block is very critical and plays an important role in data capture, processing, and usage. Substantial investments have been made in the development and strengthening of routine health information systems (RHIS) in many low- and middle-income countries (LMICs) over the past 2 decades [2,3]. Although early RHISs were produced using paper-based health facility reports, many LMICs have implemented newer web-based systems over the past decade [4,5]. Given that substantial investments have been made in strengthening RHISs in LMICs in recent years, researchers have a growing demand for more real-time data [6]. Besides, data for policy and operational decision-making in LMICs, including Nigeria, have been largely limited to report generation, process monitoring, and early surveillance responses. Reliable, quality, timely, and transparent health service data are essential for an efficient health system [7]. Globally, health care interoperability has been identified as vital to seamless care coordination among the different stakeholders.

Global Health Care Interoperability

According to the World Health Organization (WHO) Europe's 2016 e-Health in practice report, Estonia is the first country to implement electronic health records (EHRs) [8]. The famous X-Road facilitates Estonia’s exchange network, an interoperability layer launched in 2001, with several different services added over the years. Estonia achieved success with over 99% of electronic medical subscriptions in 2018. Estonia's X-Road interoperability layer connects over 2700 services across 700 institutions and enterprises across several sectors, including health care. The United Kingdom’s national service specification was based on the Health Level 7 (HL7) version 3 standard and is now transitioning to the HL7 Fast Healthcare Interoperability Resource (FHIR). However, local implementation was left to providers to determine, most of whom already run different versions of HL7 version 2 [9]. This National Health Service project started with 2 main use cases: the Summary Care Record and the Detailed Care Records [9]. Canada launched a national Infoway project to standardize and foster collaboration among pan-Canadian health care solutions [10]. After leveraging the CEN TC251 standards for referrals, discharge letters, laboratory, prescriptions, reimbursements, radiology requests, and reports, a national program was deemed successful in Denmark. These use cases were pilot-tested via 15 independently managed projects [10]. In 2008, a report highlighted that in the United States, “only 15 to 20 percent of medical doctors have access to computerized patient records and only a small fraction of the billions of medical transactions happen electronically” [10]. Such a low usage led to the creation of Health Information Technology for Economic and Clinical Health, which was launched in 2009 to incentivize digitization, and things have since changed.

Interoperability in LMICs

Some LMIC health systems services are still paper-dependent for recording and transmitting health information. Paper records are limited because only one person can access them at a time. Systematic digitization of health systems has driven the development and implementation of national digital health strategies in Nigeria and other LMICs [11,12]. From our literature search [13] and to our knowledge, LMICs still struggle with patient-level interoperability project implementations, which has limited recording successes. Nigeria is a typical LMIC because it has one of the highest global burdens of maternal mortality [14]. Furthermore, there are more primary health care (PHC) facilities (approximately 10 times) than hospitals in Nigeria; hence, here we focus on PHCs. PHCs have the highest potential for impact in the Nigerian health system because most health services are delivered at the PHC level. In a typical PHC network, the possible use cases for health information interchange may include the following:

- Interdepartmental care communication
- Inter-PHC or PHC to secondary hospital referral
- Reporting of decentralized laboratory results
- Triangulation of immunization and surveillance information
- Payment settlement
- Diagnostic information exchange

Nigeria used the DHIS2 for routine reporting of the delivery of health information system services. Routine health information systems (RHISs) continue to collect data on a wide range of diseases and conditions [6]. These RHIS data are analyzed to assess community-level initiatives such as policies to boost community engagement and strengthen referrals from traditional birth attendants to increase demand for maternal and child care [15-17]. The COVID-19 pandemic has further exposed the weakness in health systems worldwide and the value of linkages.

Health Care Interoperability Standards

The international organizations for certifying and ratifying widely used digital health standards are the ISO/TC (International Organization for Standards’ Health Informatics Technical Committee) 215 and CEN/TC (European Committee for Standards’ Health ICT Technical committee) 251. For instance, ISO 21090:2011 is a ratified HL7 version 3 data type for information interchange. Similarly, ISO 13606-1:2019 is a ratified description of archetype reference models. HL7 is a leading health care standard development organization that has facilitated many standards, including the HL7 version 2 messaging standard, HL7 version 3 Clinical Document Architecture document exchange standard, and the HL7 FHIR. FHIR was popularized because it supports REpresentational State Transfer (REST)--based web-based (real-time) transactions and its extension for services. FHIR is now emerging as the de facto global standard for health care data interchange. The FHIR community includes Microsoft, Google, Apple, and many electronic medical record and EHR vendors [18-20]. In addition, the WHO has recently published a digital adaptation kit to support countries deploying standards for antenatal care [21].
Terminologies
In addition to data interchange standards, terminology categorization helps guarantee consistent and uniform understanding (and meaning) of terms in health care systems (within and across geographies). The leading terminologies for disease, procedure, and other concept classification are Systematized Nomenclature of Medicine–Clinical Terms (SNOMED-CT) and the International Classification of Diseases (ICD). Other technology providers are Logical Observation Identifiers Names and Codes (LOINC) for laboratory result reporting and Digital Imaging and Communication in Medicine (DICOM) for imaging data reporting. This study used ICD because it uses a free license compared to the better developed SNOMED-CT for disease classification. Code systems such as the WHO ICD-10 use statistical classification of medical concepts and entities into coded groups, assigning identifiers [22]. Codes allow for the unique identification of these concepts in an information processing system. These codes classify diseases, procedures, billing, history or symptoms, and case summaries (jurisdictional and international aggregate reporting). Simultaneously, service providers, including clinicians, use clinical terms in information processing tools.

Study Objectives
The project’s main objective is to use a referral use case to profile, validate, and present data elements relevant to exchanging health information at the PHC level of care. Profiling is the strategy for defining FHIR models by domesticating the international core standard through specific use cases by structured authoring and publishing. Global best practices facilitate digital health information exchange for better care by using standardized data. Digital tools can only communicate using data in certain formats (e.g., XML or JSON), organized in an agreed structure [23].

Methods
Overview
We reviewed paper referral forms and surveyed frontline health workers, drawing inspiration from similar work conducted by Odisho et al [24]. We checked how consistent the referral data sets were. Aggregated referral data sets were then mapped to and FHIR extensions profiled. We also modeled data types and cardinalities, including references to other profiles, resources, and terminology binding to ICD-10.

Stakeholder Interviews and Data Set Identification
We established the research focus by addressing the data flow in the maternal and child health information flow value chain in Ebonyi State, Nigeria. Nigeria has between 28,000 and 36,000 health facilities overall. Ebonyi state is one of the 36 subregional governments in Nigeria with 171 “functional” PHC centers and 13 general hospitals [25]. Although from the National Health workforce Registry, there are up to 830 health facilities in the state [26]. Based on our use case, a strategic point of data exchange among multiple PHC centers or PHC centers and hospitals is the referral chain for pregnant women. We used the purposeful snowball sampling technique to identify health care providers in Ebonyi State and share the survey questionnaire.

We sent out questionnaires and a request for a copy of “referral forms” was used for 24 health workers in their respective health facilities in Ebonyi State. Between June 10 and 17, 2019, all 24 health workers completed and returned the questionnaires, and only 3 provided referral paper forms. Respondents were a mix of medical doctors, midwives, and nurses, as shown in Table 1.

We acknowledge the possibility of selection bias, and, for instance, these providers were mostly from health facilities in Abakaliki, the state capital and the main metropolitan city. We consider this bias insignificant as we measured consistency or variation in referral data sets among providers, which was significant. Each provider was from a different health facility (except the tertiary hospital doctors).

The structured questionnaire used asked the following questions:

1. What information is shared when referring-out a pregnant woman?
2. What information is expected when referring-in a pregnant woman?
3. What forms are used, and what are the contents of these forms?
4. What information is the client or caregiver expected to know or have at the recipient end?

Table 1. Distribution of respondents and their roles.

<table>
<thead>
<tr>
<th>Workstation</th>
<th>Roles, n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nurse</td>
</tr>
<tr>
<td>Primary health care clinics</td>
<td>3</td>
</tr>
<tr>
<td>Secondary health care (general hospitals)</td>
<td>4</td>
</tr>
<tr>
<td>Tertiary health care (teaching hospital)</td>
<td>0</td>
</tr>
<tr>
<td>State Ministry of Health</td>
<td>0</td>
</tr>
</tbody>
</table>

Profiling, Validation, and Publishing
We started by creating a default patient profile with no extension by using the Forge tool and uploading it on the simplifier.net web interface under the BlockMom project for validation [27]. This first step was to confirm that the example of the base patient resource instance is FHIR-conformant. From the stakeholder interviews, we aggregated information data sets. We then mapped them to the standard patient resource to create a referral resource with extensions that capture all the identified data points. We further created the bare XML schema for easy file-based resource instance validation. The codes in XML and
JSON formats are freely available on the GitHub directory [28]. Our steps and tools used are shown in Figure 1.

We modeled the FHIR referral use case profile of information flow regarding pregnant women from one PHC center—for example, PHC center 1 to PHC center 2—or general hospital. Afterward, these resource mapping outputs were then synthesized into JSON and XML machine-readable data formats on the basis of FHIR resources for antenatal referral. We have further indicated the resources category affected by our referral bundle in green in Figure 2.

**Figure 1.** Steps to profiling and publishing the Fast Healthcare Interoperability Resource.

![Diagram showing steps to profiling and publishing FHIR](image)

**Figure 2.** Resources considered for the Referral bundle.

<table>
<thead>
<tr>
<th>Foundation</th>
<th>Security</th>
<th>Conformance</th>
<th>Terminology</th>
<th>Documents</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>Individuals</td>
<td>Entities</td>
<td>Workflow</td>
<td>Management</td>
<td></td>
</tr>
<tr>
<td>Clinical</td>
<td>Clinical</td>
<td>Diagnostic</td>
<td>Medication</td>
<td>Care provision</td>
<td>Request and Response</td>
</tr>
<tr>
<td>Financial</td>
<td>Support</td>
<td>Billing</td>
<td>Payment</td>
<td>General</td>
<td></td>
</tr>
<tr>
<td>Specialized</td>
<td>Public health and research</td>
<td>Definitional artifacts</td>
<td>Clinical decision support</td>
<td>Quality reporting</td>
<td></td>
</tr>
</tbody>
</table>

**Binding With ICD-10 Terminology**

An example referral use case is described in Sierra Leone’s digital health strategy 2018-2023, “Table 2- Scenario: The vision in Practice” [12]. Asuma, a pregnant woman described in the use case, was referred to the clinic from the community by a roaming community health extension worker. See page 25 of Sierra Leone’s national digital health strategy for more information on this use case. In ICD-10, the “Personal History of malaria” code is Z8613. The code allows for unique identification in any information system using the same coding system, thus distinguishing this from, say, B500, which represents “Plasmodium falciparum malaria with cerebral complication,” which is a case of complicated malaria with intermittent coma.

We analyzed the 21 chapters of ICD-10. Table 2 highlights those that are most relevant for in-depth study when designing a Maternal and Child Health (MNCH) information management system [10].

Code O98.6 represents “Protozoal disease complicating pregnancy, childbirth, and the puerperium.” This is synonymous with “Malaria in pregnancy” or “Maternal malaria during pregnancy,” both not explicitly coded in ICD-10 [29]. “Malaria in pregnancy” is the scenario described in the Sierra Leone MNCH digital health use case from the preceding paragraph.
Table 2. Key Maternal and Child Health chapters of the International Classification of Diseases, Tenth Revision coding system.

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Description</th>
<th>Code range</th>
</tr>
</thead>
<tbody>
<tr>
<td>XV</td>
<td>Pregnancy, childbirth, and the puerperium</td>
<td>O00xx to O99xx</td>
</tr>
<tr>
<td>XVI</td>
<td>Certain conditions originating in the perinatal period</td>
<td>P00xx to P96xx</td>
</tr>
<tr>
<td>XVIII</td>
<td>Symptoms, signs, and abnormal clinical and laboratory findings not elsewhere classified</td>
<td>R00xx to R99xx</td>
</tr>
<tr>
<td>XX</td>
<td>External causes of morbidity and mortality</td>
<td>V01xx to Y98xx</td>
</tr>
<tr>
<td>XXI</td>
<td>Factors influencing health status and contact with health services</td>
<td>Z00xx to Z99xx</td>
</tr>
</tbody>
</table>

Ethics Consideration

This study was exempted from ethics approval by the University of Malta ethics review board.

Results

Survey Outputs

In addition to responses from these surveys, 3 different referral forms for referral tracking among pregnant women used in the state were made available by respondents. The forms were then mapped to survey questions to generate a unified form with a union of content from the 3 forms in Table 3.

Survey responses from care providers varied widely and included extraneous information than in the referral forms. Based on the 24 health care providers' responses for the first 2 questions—“What Information is shared when referring a pregnant woman?” and “what information is expected when receiving a pregnant woman?”—none of the responses matched for all respondents. In response to question 3—“what forms are used?”—3 respondents said “referral letter” and 7 said “referral form.” Three respondents noted that referral forms varied by health institutions, while one indicated that they do not use any forms for referral. Other forms listed by respondents are the consent form, investigation form, chemistry form, hematology form, results form, ultrasound form, laboratory form, radiology form, and virology form. In response to the question, “What information is the client, or their caregiver expected to have or know?” At the same time, 2 respondents said “nill,” the rest of them listed information that completely varied. When we mapped the aggregated responses with the form details from Table 3, the list of data sets will be similar to that shown in Textbox 1.
<table>
<thead>
<tr>
<th>Facility name</th>
<th>National Health Insurance Scheme (NHIS) referral form</th>
<th>Women, infants, and children referral form for pregnant women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facility name</td>
<td>Facility name</td>
<td>N/A(^a)</td>
</tr>
<tr>
<td>NHIS code</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td></td>
<td>Measurement date</td>
</tr>
<tr>
<td>Patient number and social insurance number</td>
<td>Health Management Organization (HMO)</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>NHIS ID number</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HMO code</td>
<td></td>
</tr>
<tr>
<td>Name of patient</td>
<td>Name</td>
<td>Patient’s name (last, first)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Date of birth</td>
<td>Date of birth (MM/DD/YY)</td>
</tr>
<tr>
<td>Sex</td>
<td>Sex</td>
<td>—</td>
</tr>
<tr>
<td>Address</td>
<td>N/A</td>
<td>Address (state, city, zip code)</td>
</tr>
<tr>
<td>Telephone number</td>
<td>N/A</td>
<td>Telephone number</td>
</tr>
<tr>
<td>Complaints</td>
<td>Presenting complaint</td>
<td>—</td>
</tr>
<tr>
<td>Findings on examination</td>
<td>Examination findings</td>
<td>Height, weight, Hemoglobin (g/dL), hematocrit (%), and blood test date</td>
</tr>
<tr>
<td>Investigations performed, if any</td>
<td>Investigation results</td>
<td>—</td>
</tr>
<tr>
<td>Provisional diagnosis</td>
<td>Provisional diagnosis</td>
<td>—</td>
</tr>
<tr>
<td>N/A</td>
<td>Reason for referral post medical history taking</td>
<td>Estimated date of confinement</td>
</tr>
<tr>
<td></td>
<td>Clinical warnings (allergies)</td>
<td>Date when last pregnancy ended</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gravida Para</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pregravid weight (lbs)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Indicate any of the following medical conditions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(diabetes, multiple pregnancies, hypertension, tuberculosis,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>previous poor pregnancy outcome, and history (specify)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>If other, current history of the condition (specify)</td>
</tr>
<tr>
<td>Current and recent medication</td>
<td>N/A</td>
<td>Current medication and supplements prescribed</td>
</tr>
<tr>
<td>N/A</td>
<td>Other relevant information</td>
<td>Impressions and comments</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Comments</td>
</tr>
<tr>
<td>Name of officer</td>
<td>Referring doctor</td>
<td>Name of the physician care provider group and clinic</td>
</tr>
<tr>
<td></td>
<td>Medical and Dental Council of Nigeria number</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Receiving doctor’s Medical and Dental Council of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nigeria number</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Date</td>
<td></td>
</tr>
<tr>
<td>Designation</td>
<td>N/A</td>
<td>—</td>
</tr>
<tr>
<td>Signature</td>
<td>Signature and stamp</td>
<td>Health care provider</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Signature Date</td>
</tr>
<tr>
<td>To</td>
<td>Health facility</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>NHIS code</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)N/A: not applicable.
**Textbox 1.** List of data sets.

- Source health facility name
- Source health facility ID
- Destination health facility name
- Destination health facility ID
- Date of referral
- Patient name
- Patient number
- Patient no type (Health Management Organization, National Health Insurance Scheme, softwareVendor)
- Patient age
- Sex
- Address
- Complaint
- Presenting complaint
- Investigation done
- Findings on examination
- Provisional diagnosis
- Reason for referral
- Current medication
- Recent medication
- Name of referring officer
- ID of referring officer (Medical and Dental Council of Nigeria)
- Designation of the referring officer
- Other relevant information
- Referrals direction (in or out)
- Referrals by disease
- Malaria case referred for adverse drug reaction (Health Management Information System)
- Referral disease (tracked by age and case)

**Referral FHIR Resource (Known as Referral Letter or Discharge Letter)**

The profile developed in this paper is only considered a provisional national profile suggestion for consideration and should not be relied on for clinical decisions. The resource mapping's final output is in XML and JSON formats and is freely available on GitHub [28]. The resource file is shown in Figure 3.
REST and FHIR

Our artifact assumes the use of the REST paradigm for information exchange. In this section, we explain the technicalities of REST. REST is the foundation for the scale of the internet as we know it today [30]. While FHIR supports many different communication paradigms, REST is responsible for its popularity. The REST paradigm leverages the HTTP protocol with the simple client-server architecture with variants of catch-less/catching, stateless/stateful, or n-tiered architectures and hierarchies. For further technical details of these ICT concepts, please refer to Thomas Fielding’s thesis introducing REST in 2000 [30].

Similar to traditional REST, FHIR’s REST paradigm considers entities and concepts as resources. Each of the resource instances is unique and is represented using a uniform resource identifier (URI). The URI may also be used for locating the resource if it points to the location on a given server (in which case, it can also be referred to as URL). There are a finite number of ways a client can manipulate entities and concepts (resources) located in an FHIR server using REST requests of get, put, post, delete, options, head, trace, and connect.

The server responds to the client’s HTTP requests after performing internal business logics unique to each server implementation. The clients will receive the same response for similar HTTP requests irrespective of whether they are for a mobile app, web browser, computer application, or embedded device, as shown in Figure 4.

Both the client and server use the header and body components of the HTTP request or response for their information exchange, depending on the HTTP method used. The server always returns a status code indicating success or failure or a variant of either with further detail for each request. FHIR servers can use the OperationOutcome resource to provide structured details of request failure to the clients in the event of failure. There are over 100 different FHIR resources [31]. When the request succeeds, the client is sent the resource by the server.

Methods may or may not be allowed (or even implemented) by the server for a particular resource and may be specified by the client’s server response. The header has many attributes that can be set, for instance, to indicate the data type it accepts, authorization credentials, connection, content encoding, caching, and more. The content-type attribute can indicate the resource as either XML or JSON format—both native to FHIR. FHIR, similar to REST, is an open standard and thus aligns with key principles of digital development [32].

Figure 3. The profiled referral resource in picture.
Discussion

Principal Findings

We could not use the core FHIR resource because it did not contain all the data sets as aggregated in Figure 4. To add and extend the outstanding data sets, FHIR designers provide for and allow extendibility using profile extensions. FHIR contributors and balloting-process use the 80-20 rule to determine what makes it into the core FHIR resource [33]. It is understandable if there are no contributions from Nigeria or many other LMICs because they are often not represented at HL7 FHIR balloting. Our referral form reviews and data set mapping also lay credence to our hypothesis that the core patient resource needs an extension for our use case.

Our study shows that traditional paper referral forms currently in use vary widely in Nigeria’s PHC system as illustrated in the word art in Multimedia Appendix 1. The implication is that interinstitution care coordination will remain suboptimal as much of the essential information will be missing. This work will help policy makers and PHC centers in Nigeria understand the need to standardize or enforce agreed referral standards. In addition, the steps we have outlined in this work will help guide institutions as they standardize or adopt FHIR.

While the surveys were conducted before the COVID-19 pandemic, their findings are relevant for continued functioning of the PHC centers even amid and after the COVID-19 pandemic. The pandemic has exposed the weakness of health systems and shown the importance of interconnected and interoperable health systems. Emerging technologies are being proposed in response to the pandemic [34] and new models are emerging for health information interoperability [35] in LMICs. Even when PHC centers are digitized, referrals among health facilities in many LMICs with different software vendors do not happen seamlessly. Our research shows that referral practices for pregnant women varied significantly even in urban settings.

The key output of this study is the FHIR referral resource artifact, which will help vendors design consistent referral data sets and ensure out-of-the-box interoperability. This resource remains broad and from the core patient resource. FHIR allows for organizational or national extensions and adaptations [33]. Health authorities in many LMIC countries will benefit from standardizing and exposing its required referral data model for Women and Child Health, which encompasses MNCH [24]. Publishing a public FHIR specification that can be leveraged by MNCH solution implementers will help simplify interface implementations.

We here illustrate that our approach differs from traditional health information exchange approaches. Figure 5A shows the traditional document- and message-based information (which is still supported by FHIR) where both databases are required to retain the messages being exchanged. Figure 5B shows that the end point query approach using REST method calls is used to access or share FHIR resources being exchanged.

Women continue to die owing to preventable causes at the point of giving birth. Many of these deaths happen before, during, and after delivery. In addition, maternal health has been highlighted in some LMIC national digital health strategies (ie, Nigeria and Sierra Leone) as a priority health area [11]. Furthermore, our BlockMom model used the ICD terminology over the SNOMED-CT model owing to its favorable pricing license. For instance, SNOMED licenses are based on the number of health facilities using the terminology service, though there are requirements to apply for a waiver for certain implementations in LMICs. Moreover, the model’s deliberate focus on FHIR over other HL7 or ISO standards was because it is free and open for adaptation, adoption, and testing in LMICs.
Figure 5. Fast Healthcare Interoperability Resource (FHIR) paradigms in a 2–public health care (PHC) system referral exchange scenario. EHR: electronic health record. REST: REpresentational State Transfer.

Limitations
We used on-file schema and the ICD-10 version 2021 text file to validate profiled resources. Even though the bare schema was provided in the GitHub directory, standard practice would be to set up an FHIR server for this purpose. This is the focus of our future study. Although many mobile-based solutions are available in PHC centers in LMICs, these profiles can only be used with a mobile-based solution that uses patient-specific information rather than aggregated information, as is the case with many community health worker information systems. Furthermore, ICD-11 2022 release has just been launched by the WHO, which emphasizes following the profiling process and not necessarily the output [36].

We also noted that unique identification metrics and characteristics, as explained by McFarlane et al [37] and Chukwu [38], were not part of the data sets proposed and made available during referral. This aspect was profiled as part of the FHIR resource. Besides, this may be the case as the patient-required abridged historical information is assumed to be comprehensive enough. Nevertheless, in a digital platform where a unique identifier is important, this assumption will not hold true. We assumed cryptographically generated unique identification mechanisms in this prototype.

A key limitation of our survey approach was that we did not prevalidate the questionnaire before use. In addition, our sampling methodology used a snowball strategy that targeted health workers from the most urban part of Ebonyi State, Nigeria. We are aware and acknowledge that this may seem inherently biased. However, our aim for the survey was to determine consistency or otherwise of the referral data sets, which we determine to vary widely across all respondents. In addition, since Abakaliki is the state capital and the main metropolitan city in the state, it is expected to have a standardized referral form; however, it does not.

Conclusions
Questionnaire responses were collected from health care providers, and referral forms from health institutions in Nigeria were reviewed. Survey responses and fields of referral forms show variability in referral data sets across respondents and forms. Here we have made a case for FHIR, an emerging health care data interchange standard, and have profiled a referral resource for PHC information exchange targeted at LMICs. This paper describes the profiling steps, including key questionnaire responses and mapping of referral forms. We have proposed the use of ICD-10 terminology and used file-based schema validation. The methodology and artifacts will be invaluable for the research and implementation community targeting LMICs. Our future work will set up the server and configure the appropriate binding for this and other resources.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Survey respondents’ referral keywords.
[ PNG File , 2318 KB - formative_v6i7e28510_app1.png ]

References


https://formative.jmir.org/2022/7/e28510

JMIR Form Res 2022 | vol. 6 | iss. 7 | e28510 | p.248

(page number not for citation purposes)
31. HL7 FHIR. URL: http://hl7.org/fhir [accessed 2021-01-10]
33. FHIR and Architectural Principles. HL7 FHIR. URL: https://build.fhir.org/overview-arch.html#-text=In%20addition%2C%20FHIR%20aligns%20to,80%25%20of%20the%20interoperability%20needs [accessed 2022-06-13]

Abbreviations

- CEN: European Committee for Standards’ Health ICT
- DICOM: Digital Imaging and Communication in Medicine
- EHR: electronic health record
- FHIR: Fast Healthcare Interoperability Resource
- ICD: International Classification of Diseases
- ICT: Information and Communication Technologies
- ISO: International Organization for Standards
- LMIC: low- and middle-income country
- LOINC: Logical Observation Identifiers Names and Codes
- MNCH: Maternal and Child Health
- PHC: primary health care
- REST: REpresentational State Transfer
- RHIS: routine health information systems
- SNOMED-CT: Systematized Nomenclature of Medicine–Clinical Terms
- URI: uniform resource identifier
- WHO: World Health Organization

Edited by A Mavragani; submitted 05.03.21; peer-reviewed by J Ye, B Nievas Soriano; comments to author 14.05.21; revised version received 23.05.21; accepted 14.03.22; published 07.07.22.

Please cite as:
URL: https://formative.jmir.org/2022/7/e28510
doi: 10.2196/28510
PMID:35797096

©Emeka Chukwu, Lalit Garg, Nkiruka Obande-Ogbuinya, Vijay Kumar Chatta. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.07.2022. 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.
Understanding the Security and Privacy Concerns About the Use of Identifiable Health Data in the Context of the COVID-19 Pandemic: Survey Study of Public Attitudes Toward COVID-19 and Data-Sharing

Charlotte Summers1, BSc; Frances Griffiths2, PhD; Jonathan Cave2,3,4, PhD; Arjun Panesar1, MEng

1DDM Health, Coventry, United Kingdom
2Warwick Medical School, University of Warwick, Coventry, United Kingdom
3Department of Economics, University of Warwick, Coventry, United Kingdom
4Data Ethics Group, The Alan Turing Institute, London, United Kingdom

Corresponding Author:
Charlotte Summers, BSc
DDM Health
Technology House, Science Park
University of Warwick
Coventry, CV4 7EZ
United Kingdom
Phone: 44 7969091134
Email: charlotte@ddm.health

Abstract

Background: The COVID-19 pandemic increased the availability and use of population and individual health data to optimize tracking and analysis of the spread of the virus. Many health care services have had to rapidly digitalize in order to maintain the continuity of care provision. Data collection and dissemination have provided critical support for defending against the spread of the virus since the beginning of the pandemic; however, little is known about public perceptions of and attitudes toward the use, privacy, and security of data.

Objective: The goal of this study is to better understand people’s willingness to share data in the context of the COVID-19 pandemic.

Methods: A web-based survey was conducted on individuals’ use of and attitudes toward health data for individuals aged 18 years and older, and in particular, with a reported diagnosis of a chronic health condition placing them at the highest risk of severe COVID-19.

Results: In total, 4764 individuals responded to this web-based survey, of whom 4674 (98.1%) reported a medical diagnosis of at least 1 health condition (3 per person on average), with type 2 diabetes (n=2974, 62.7%), hypertension (n=2147, 45.2%), and type 1 diabetes (n=1299, 27.4%) being most prominent in our sample. In general, more people are comfortable with sharing anonymized data than personally identifiable data. People reported feeling comfortable sharing data that were able to benefit others; 66% (3121 respondents) would share personal identifiable data if its primary purpose was deemed beneficial for the health of others. Almost two-thirds (n=3026; 63.9%) would consent to sharing personal, sensitive health data with government or health authority organizations. Conversely, over a quarter of respondents (n=1297, 27.8%) stated that they did not trust any organization to protect their data, and 54% (n=2528) of them reported concerns about the implications of sharing personal information. Almost two-thirds (n=3054, 65%) of respondents were concerned about the provisions of appropriate legislation that seeks to prevent data misuse and hold organizations accountable in the case of data misuse.

Conclusions: Although our survey focused mainly on the views of those living with chronic health conditions, the results indicate that data sensitivity is highly contextual. More people are more comfortable with sharing anonymized data rather than personally identifiable data. Willingness to share data also depended on the receiving body, highlighting trust as a key theme, in particular who may have access to shared personal health data and how they may be used in the future. The nascency of legal guidance in this area suggests a need for humanitarian guidelines for data responsibility during disaster relief operations such as pandemics and for involving the public in their development.
Introduction

The World Health Organization declared the COVID-19 outbreak a public health emergency on January 30, 2020; after 6 weeks, it was categorized as a pandemic [1]. Certain groups of people are particularly likely to have serious or severe symptoms of COVID-19 [2]. Preliminary data suggest that people with obesity are at an increased risk of severe COVID-19 [3]. Type 2 diabetes mellitus and hypertension are the most common comorbidities in patients with COVID-19 [4]. According to several reports, including those from the Centers for Disease Control and Prevention, patients with type 2 diabetes are at a greater risk of death than those without type 2 diabetes [5].

Digital health technologies are being used in the fight against COVID-19 [6]. Global health care systems have seen an influx in the incidence of the same novel condition, and the contagious nature of the condition has driven the shift to remote medicine. Many health care pathways have been rapidly digitalized with face-to-face services seeing a drop in usage [7]. This has increased the collection, sharing, and use of data in digital form. Technology is used for remote monitoring, general practitioner consultations, providing structured education, and tracking the spread of disease. As well as the technologies themselves, the data they generate are also useful [8].

Timely, secure, and reliable data access and sharing are critical to understanding COVID-19, controlling its spread, improving the effectiveness and acceptance of government policies, and fostering global cooperation in the race to develop and distribute effective therapies and vaccines. During the COVID-19 pandemic, data are being rapidly shared to understand the location of infections, confirmed cases, recoveries, and deaths. The main data points of interest for this are geolocation and biometric data, both of which are available from users’ mobile devices. However, there are serious concerns regarding the objectivity and accuracy of these data, and their utility has been compromised by inconsistent collection and definitions. This, in turn, feeds back into individuals’ trust in the collecting organizations and in the extent to which their shared data will actually be used to help others, and this needs to be matched by the trustworthiness of those organizations.

During an unprecedented time, some digital responses to the crisis have precipitated novel data governance and privacy challenges [9]. Governments are taking extraordinary measures to track, trace, and contain the spread of COVID-19 by transitioning to digital technologies and advanced analytics to collect, process, and share data for effective frontline responses. Government-mandated apps are bringing the fight against COVID-19 onto users’ devices and have generally adopted pragmatic and contextualized approaches, but they have prompted concerns about security and privacy and the control and use of data beyond the pandemic [10]. There is a trade-off between effectiveness and privacy, centralized and decentralized implementations, and the links to trace and isolate policies.

While the exceptional measures implemented in some countries may prove effective in limiting the spread of the virus, some have provoked controversy in terms of privacy and other fundamental rights, particularly when they lack transparency and public consultation [11]. In South Korea, the specificity of publicly available anonymized data raised privacy concerns when some researchers found that data trails were so detailed that individuals could be identified [12]. In Italy, the Department of Prevention released specific guidelines on the application of the European Union’s and national data privacy rules in the context of the COVID-19 pandemic [13]. Similarly, the United Kingdom’s Information Commissioner’s Office, an independent authority set up to uphold information rights in public interest, confirmed that there would be no regulatory action taken against organizations that fail to meet the data protection standards if noncompliance results from the COVID-19 pandemic [14,15]. In China, new arrivals to the country are tested for COVID-19, instructed to download a government-mandated app, and wear a wristband that is linked to the app to monitor movement with a technology similar to that used in Singapore [16,17]. The United Kingdom’s Track and Trace app was the center of a debate on centralization of data [18]. On May 5, 2020, the Government revealed its first attempt at a contact-tracing app, but 6 weeks later admitted that the app was flawed and it would switch to a more privacy-preserving model devised by Apple and Google [19,20]. Transparency is a key theme. One of the most common misconceptions about the United Kingdom’s Track and Trace app was that it could allow users to specifically identify and map COVID-19 cases among their contacts and in their vicinity [21].

This study seeks to understand the opinions of British people with long-term health conditions on the themes of data privacy and security, data ethics, and data misuse and to assess the possible trade-offs in data utilization to manage a crisis such as the COVID-19 pandemic [22]. It is important to understand the concerns of people with long-term health conditions such as type 2 diabetes and hypertension as these conditions have been shown to be key risk factors in the progression and prognosis of COVID-19 [23,24].

Methods

Study Design and Setting

A web-based survey study was conducted with a mixed methods design conforming to the checklist for reporting results of internet electronic surveys [25]. An email invitation to participate, which included a weblink to the survey, was sent to 11,213 people who had consented to be contacted for research opportunities.
Quantitative information (closed and multiple-choice questions) was collected on four topics: (1) demographic characteristics, (2) COVID-19 symptoms and clinical diagnoses, (3) sharing and privacy of pre- and post–COVID-19 health data, and (4) COVID-19 lockdown behaviors. Responses from the final topic are not included in this analysis.

The survey contained 31 questions: 26 closed questions, 1 open question, and 4 demographic questions. Questions on sharing and privacy of pre- and post–COVID-19 health data were answered on a 5-point Likert scale with responses ranging from strongly disagree to strongly agree or from not concerned at all to very concerned.

Participants
People aged ≥18 years who had joined the Diabetes.co.uk community were surveyed. The survey commenced with 1 screening question: “Do you consent to take part in the study?” Respondents who consented went on to complete the survey.

Procedure
Data collection occurred between July 6 and August 31, 2020. The survey was administered through the Jisc Online Surveys software and comprised closed, open, and multiple-choice questions. The survey was designed to elicit individual responses to questions about retrospective data use and privacy prior to the COVID-19 pandemic and prospective use during the COVID-19 pandemic.

It is intended to have multiple windows of data collection for several reasons: people’s recollections of pre–COVID-19 attitudes may be unreliable, and changes in the course, apparent seriousness, and confidence in scientific understanding of the pandemic will have evolved.

The type and wording of each question was composed by the research team. The order of questions was not randomized. The survey followed a predetermined logic where contingent questions were included or automatically skipped on the basis of responses. Qualitative data were collected with 1 open question exploring what respondents would like to see happening: “What would you like to see happen to improve the COVID-19 situation?” (question 30).

Analysis
We exported all data from Jisc and conducted data analysis using SPSS (version 22; IBM Corp). We conducted descriptive data analyses of sample distributions and characteristics. Pearson r correlation coefficients were used to determine the relation between prior data-sharing behavior and attitudes toward data-sharing activity in the context of the COVID-19 pandemic. The data from the open question were read through and then categorized into themes.

Ethical Considerations
Ethics approval was obtained from the Human Research Ethics Committee of the University of Warwick (BSREC 144/19-20). Web-based informed consent was required before the survey could be accessed.

Results
Survey Respondents
Of 11,213 people emailed, 10,705 clicked through to the survey; in total, 4764 gave their consent and began the survey. As indicated in Table 1, all of them completed the survey and were included in the analysis. All respondents were located in the United Kingdom. In total, 2287 (48.0%) respondents were male and 3083 (64.8%) were aged between 55 and 74 years. A total of 115 (2.8%) respondents reported having been clinically diagnosed with COVID-19. The majority of patients (n=4674, 98.1%) reported a prior clinical diagnosis of at least one health condition (on average 3 per person). There was a high prevalence of individuals living with type 2 diabetes (n=2974, 62.7%), hypertension (n=2147, 45.2%), type 1 diabetes (n=1299, 27.4), obesity (n=892, 18.8%), and depression (n=871, 18.3%). Respondent demographics are shown in Table 1.
Table 1. Respondent demographics (N=4764).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Respondents, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2287 (48.0)</td>
</tr>
<tr>
<td>Female</td>
<td>2435 (51.1)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>42 (0.9)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>18 to 24</td>
<td>23 (0.5)</td>
</tr>
<tr>
<td>25 to 34</td>
<td>104 (2.2)</td>
</tr>
<tr>
<td>35 to 44</td>
<td>298 (6.3)</td>
</tr>
<tr>
<td>45 to 54</td>
<td>839 (17.6)</td>
</tr>
<tr>
<td>55 to 64</td>
<td>1550 (32.6)</td>
</tr>
<tr>
<td>65 to 74</td>
<td>1533 (32.2)</td>
</tr>
<tr>
<td>75 or older</td>
<td>410 (8.6)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>7 (0.1)</td>
</tr>
<tr>
<td><strong>Health conditions</strong></td>
<td></td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>2974 (62.7)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>2147 (45.2)</td>
</tr>
<tr>
<td>Type 1 diabetes</td>
<td>1299 (27.4)</td>
</tr>
<tr>
<td>Arthritis</td>
<td>1002 (21.1)</td>
</tr>
<tr>
<td>Obesity</td>
<td>892 (18.8)</td>
</tr>
<tr>
<td>Depression</td>
<td>871 (18.3)</td>
</tr>
<tr>
<td><strong>Employment</strong></td>
<td></td>
</tr>
<tr>
<td>Full-time employment</td>
<td>1205 (24.9)</td>
</tr>
<tr>
<td>Part-time employment</td>
<td>571 (11.8)</td>
</tr>
<tr>
<td>Retired</td>
<td>2298 (47.5)</td>
</tr>
<tr>
<td>Student</td>
<td>39 (0.8)</td>
</tr>
<tr>
<td>Unemployment</td>
<td>464 (9.6)</td>
</tr>
<tr>
<td>Furloughed</td>
<td>188 (3.9)</td>
</tr>
<tr>
<td>Volunteering in my community (National Health Service, key services)</td>
<td>69 (1.4)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Indian or Pakistani</td>
<td>69 (1.4)</td>
</tr>
<tr>
<td>Black, British African, or Caribbean</td>
<td>46 (1.0)</td>
</tr>
<tr>
<td>Middle Eastern</td>
<td>5 (0.1)</td>
</tr>
<tr>
<td>Mixed groups</td>
<td>27 (0.6)</td>
</tr>
<tr>
<td>White</td>
<td>4,434 (93.1)</td>
</tr>
<tr>
<td>Other</td>
<td>44 (0.9)</td>
</tr>
<tr>
<td>Chinese, Japanese, or East Asian</td>
<td>7 (0.1)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>132 (2.8)</td>
</tr>
</tbody>
</table>

*aFrequently occurring health conditions selected.

*bRespondents selected multiple statuses; for example, full-time employed and furloughed.
COVID-19 Symptoms and Clinical Diagnosis

Of the 4764 respondents who completed the survey, 494 stated they had had symptoms of COVID-19. The most common symptoms were the following: a continuous cough (n=467, 94%), fever (n=325, 65.7%), difficulty breathing (n=384, 77.7%), and loss of taste (n=324, 65.5%). In total, 384 (77.7%) respondents reported another symptom, predominantly fatigue (16.4%).

Of those reporting symptoms, 111 (22.5%) reported a clinical diagnosis of COVID-19. Of these respondents, 73 (63.5%) reported that their symptoms were severe or very severe. In total, 26 (22.6%) respondents reported that their symptoms were not severe at all. All respondents who reported a clinical diagnosis of COVID-19 reported at least one symptom, including loss of smell or taste (63.1%), fever (62.2%), difficulty breathing (61.3%), or continuous cough (53.2%).

A total of 131 (2.8%) respondents reported that a household member had been tested and was clinically diagnosed with COVID-19.

Sharing and Privacy of Pre–COVID-19 Health Data

Prior to the COVID-19 pandemic, almost half of the respondents (n=2313, 49.2%) agreed or strongly agreed that they often consented to anonymized sharing of their private health data, while only 608 (13%) respondents often consented to sharing of private health data without anonymization. Two-thirds of respondents (n=3113, 66.7%) disagreed or strongly disagreed with sharing their private health data without anonymization. Similarly, 3121 (66.3%) respondents would share their data if it keeps other people healthy; 3026 (63.9%) respondents agreed or strongly agreed to sharing private health data with the government or health authority; 1911 (40.7%) respondents agreed or strongly agreed to share their private health data with services that provide health services to the National Health Service (NHS) such as the Low Carb Program and PushDoctor. Only 232 (5%) participants agreed or strongly agreed to share private health data with social media platforms. Over a quarter of respondents (n=1297, 27.8%) agreed or strongly agreed that they did not trust any organization to protect their private health data. Just under a quarter of respondents (n=1094, 23.5%) agreed or strongly agreed that they were not concerned by the implications of sharing private health data. General health data–sharing responses are shown in Table 2. Respondents who reported that they felt “neutral” in response to the statements were excluded.

Table 2. General health data sharing responses.

<table>
<thead>
<tr>
<th>Question</th>
<th>Disagree or strongly disagree, n (%)</th>
<th>Agree or strongly agree, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I often consent to share my private health data with any organisation as long as it is anonymised</td>
<td>1273 (27.1)</td>
<td>2313 (49.2)</td>
</tr>
<tr>
<td>I often consent to share my private health data to any organisation without anonymisation</td>
<td>3113 (66.7)</td>
<td>599 (13)</td>
</tr>
<tr>
<td>I am not concerned about the implications of sharing my private health data</td>
<td>2528 (54.2)</td>
<td>1094 (23.5)</td>
</tr>
<tr>
<td>I don't trust any organisation to protect my private health data</td>
<td>1541 (33.8)</td>
<td>1297 (27.8)</td>
</tr>
<tr>
<td>I'm happy to share my private health data if it helps keep other people healthy</td>
<td>526 (11.2)</td>
<td>3121 (66.3)</td>
</tr>
</tbody>
</table>

Sharing and Privacy of Post–COVID-19 Health Data

Over half (n=3026, 63.9%) agreed or strongly agreed to share their private data with the government or health authority if asked; 1911 (40.7%) respondents would happily consent to share their private data with services that provide health services to the NHS such as the Low Carb Program and PushDoctor, if asked. Only 232 (5%) participants agreed or strongly agreed that they would consent to sharing private data with social media if asked.

Almost half of respondents (n=2228, 47.1%) were concerned or very concerned about who would have access to their personal health data in the context of the COVID-19 pandemic and 2310 (49.1%) respondents were concerned or very concerned about how their personal health data may be used in the future. Almost two-thirds of respondents (n=3054, 65%) were concerned or very concerned around the legislation of data misuse.

Just over a third of respondents (n=1563, 33.4%) would consent to share their private data with any organization if it was providing essential COVID-19 support services such as the supermarkets, pharmacies, and banks. Responses toward the use of post–COVID-19 patient data is shown in Table 3, along with the sentiment toward the use of patient data in the context of the COVID-19 pandemic, and Table 4 shows the sentiment toward future use or misuse of data collected and used under the provisions of the COVID-19 pandemic.
Table 3. Responses toward the use of post–COVID-19 patient data.

<table>
<thead>
<tr>
<th>Question</th>
<th>Disagree or strongly disagree</th>
<th>Agree or strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I would happily consent to share my private health data with the government or health authority</td>
<td>728 (15.4)</td>
<td>3026 (63.9)</td>
</tr>
<tr>
<td>I would happily consent to share my private health data with social media e.g Twitter, Facebook, Google</td>
<td>4023 (85.9)</td>
<td>232 (5)</td>
</tr>
<tr>
<td>I would happily consent to share my private health data with services that provide health services to the NHS such as Low Carb Program, PushDoctor, Babylon Health</td>
<td>1351 (18.8)</td>
<td>1911 (40.7)</td>
</tr>
</tbody>
</table>

*NHS: National Health Service.

Table 4. Sentiment toward future use or misuse of data collected and used under the provisions of the COVID-19 pandemic.

<table>
<thead>
<tr>
<th>Question</th>
<th>Not concerned at all</th>
<th>Concerned or very concerned</th>
</tr>
</thead>
<tbody>
<tr>
<td>In light of COVID-19, how concerned are you about who would have access to your personal health data?</td>
<td>1138 (14)</td>
<td>2228 (47.1)</td>
</tr>
<tr>
<td>How concerned are you about how your personal health data may be used in the future?</td>
<td>1162 (24.7)</td>
<td>2310 (49.1)</td>
</tr>
<tr>
<td>How concerned are you around the legislation of data misuse?</td>
<td>644 (13.6)</td>
<td>3054 (65)</td>
</tr>
</tbody>
</table>

Prior Willingness to Share Data

Correlations between retrospective data-sharing that happened in the context of generalized concerns and attitude changes associated with the course of the pandemic were determined. Changes were not linked to any specific studies, policies, or measures. There were strong correlations in the attitudes of people exhibiting high levels of concern about future uses of shared data and concerns about access ($r_{4685}=0.816; P<.001$). There was a strong correlation between people exhibiting attitudes of concern that firmer legislation for data misuse is needed and concerns about future repurposing and reuse of personal health data collected during the COVID-19 pandemic ($r_{4663}=0.636; P<.001$). The Pearson $r$ correlation coefficient as a normalized measure of the strength of a possible linear correlation, lying between $-1$ and $+1$. The Pearson $r$ correlation coefficient measures nonlinear correlations (eg, when extreme views are highly correlated but more moderate ones are more independent).

Respondents agreed to share their personal data with roughly the same parties prior to the COVID-19 pandemic and within the context of the COVID-19 pandemic; governments and health authorities ($r_{4710}=0.762; P<.001$), health service providers such as the Low Carb Program and PushDoctor ($r_{4662}=0.783; P<.001$), and social media platforms such as Twitter, Facebook, and Google ($r_{4662}=0.736; P<.001$).

COVID-19 News and Information

Of the 4764 respondents, 2666 (56.1%) were concerned that they may be receiving misinformation about COVID-19 from trusted sources, 1079 (22.7%) were not concerned (genuinely unconcerned and those who feel that they are in control of the consumption of news and information), and 1006 (21.2%) had never considered it.

In total, 4237 responded to the open-ended question of what they would like to see happening to improve the COVID-19 situation. The majority of respondents shared a single response: 1348 (31.8%) stated they would like to see a reliable vaccine and treatment, 884 (20.8%) stated they would like to see balanced information from the government, and 485 (11.4%) wanted to see stricter measures to prevent the transmission of COVID-19.

Discussion

Principal Findings

Our study provides insights into public perception and attitudes toward the use of identifiable health data in the context of the COVID-19 pandemic; in particular, the perspectives of those living with chronic, long-term health conditions, with an average of 4 health conditions reported per respondent. Our study suggests that data sensitivity is highly contextual. A significant proportion of people felt that their own attitudes had shifted as a result of the COVID-19 pandemic. More people reported being comfortable with sharing private health data with any organization during rather than before the COVID-19 pandemic. In order, people appear to trust their data with the government, health organizations, and social media. There is significant distrust of private health data use by social media organizations (eg, Twitter, Facebook, and Google) even though social media is used as a channel for communication by people caught up in crises such as emergency relief operations after earthquakes, tsunamis, and typhoons; where it provides a trusted and highly salient source of information about what is happening and what to do [26,27]. This is surprising as although users worldwide report that privacy and use of personal data are important issues, most rarely make an effort actively to protect these data and often even give them away voluntarily on social media where even innocuous data can reveal sensitive health information when suitably processed [28,29]. People treat data revelation and sharing differently depending on the perceived sensitivity of the data, and the sensitivity attached to different types of data is neither stable nor uniform.

When examining the correlations between retrospective views of data-sharing behavior and comfort regarding data-sharing in
the context of the COVID-19 pandemic, individuals were
comfortable (or not) in sharing personal data with the same
organizations prior to and during the COVID-19 pandemic,
suggesting that COVID-19 has not drastically shifted people’s
willingness to share or withhold their personal data. This may be
because attitudes have shifted both retrospectively and
prospectively and also depends on whether people accurately
remember and report their past views and actions. One of the
strongest correlations observed in the analysis was between
high levels of concern about the requirement for stronger
legislation protecting individuals from data misuse and future
repurposing and reuse. This highlights the need for improved
communication, transparency, and potentially stronger regulation
on how such data may be repurposed in the future, who will be
accountable for inappropriate use of data, and a commitment
to cease or reverse exceptional uses of data when the crisis is
over. Individuals’ data rights are protected by law in regulation
such as General Data Protection Regulation 2018 in Europe and
Health Insurance Portability and Accountability Act in the
United States, which make clear the scope, purpose, and time
limitations of data usage [30,31]. Concerns may therefore reflect
ignorance of existing rules, doubts over enforcement, or a belief
that current legislation does not go far enough (for instance, in
the requirement of erasure after 3 years rather than a shorter
time duration).

A key theme emerging from the literature that was confirmed
in this study is the importance of trust [32-34]. Over a quarter
of respondents stated they did not trust any organization to
protect their data, over half reported concern about the
implications of sharing personal information, and almost
two-thirds were concerned about data misuse regulation not
being strict enough. When asked during the pandemic (the
United Kingdom’s first wave), almost half of respondents were
concerned about who would have access to their personal health
data and a similar number were concerned about how their
personal health data might be used in the future. This is
consistent with prior research suggesting that public involvement
in data policy is crucial to bolstering trust and provides support
for legislation that is more enforceable [35]. Attitudes may have
been perturbed by news stories relating to cybersecurity and
privacy and by policy announcements (eg, around Huawei, the
Online Harms Bill, etc) [36,37].

Although there are no directly comparable studies, the results
from this study complement prior research on public perceptions
about COVID-19 and data-sharing. Data privacy and
protection are important concepts [38]. Data policy tends to
address human concerns about privacy by making rules about
data protection; however, this can lead to category errors since
data protection can undermine privacy.

Willingness to share anonymized personal health information
varies depending on the degree to which the receiving body is
trusted and the uses to which the data will be put [39,40]. The
more commercial the objectives of the receiving institution
appear, the less respondents are willing to share their personal
health information. This in turn suggests that anonymization’s
disadvantages (in terms of confirming data and correlating
shared with other data) might be offset by better (wider, deeper,
and more accurate) sampling leading to greater validity of
results. Further evidence comes from the interaction (or
correlation) between these attitudinal responses and other
characteristics, meaning that nonanonymized collection might
lead to biased results.

Virus tracking apps are used at scale by governments; however,
concerns about transparency, privacy, and morality remain
[41,42]. There has been substantial research into the challenges
involved in the digital response to the COVID-19 pandemic
and proposed methodologies for the ethical design and use of
digital public health tools [43,44]. Clear and effective data ethics
is both a moral and a practical obligation. The nascentness of legal
guidance in this area combining ethics, law, and humanitarian
impulses suggests the requirement for humanitarian guidelines
for data responsibility during global crises such as pandemics.
Therefore, rather than recalibrating the expectations of people
with regard to their own privacy, the requirements for the use
of data should be broader and more comprehensive as ethically
collected big data could prove to be extremely useful in the
prediction, monitoring, and mitigation of pandemics such as
COVID-19 [45].

Strengths and Limitations

Despite the importance of the findings reported here, it is
important to note that this study had several limitations. Conducting this study via a web-based survey carries a risk of
response bias, simply because the respondents are likely to be
more technology-savvy than the general population. However,
the population studied (those with chronic health conditions) is
of interest as these participants have a degree of awareness and
the ability to self-manage their condition that is not (yet) typical
of the population at large, and this sheds light on how policies
that raise awareness may lead to greater effectiveness in terms
of uptake of technical solutions and effectiveness of public
health advice and other policies.

Participants were asked to rate retrospectively their perceptions
of data sharing prior to the COVID-19 pandemic; these ratings
may be inaccurate owing to faulty memories and response bias.
There is some ambiguity between what people (now) thought
they would have done had they been asked and how they
responded to actual requests for consent to data-sharing. In
particular, one could disagree with a statement like “I often
consent to share my private health data to any organisation
without anonymisation” simply because one rarely recalls being
asked to share (even without a principled objection to such
sharing should the occasion arise). In addition, the phraseology
of the questionnaire refers to data that have either been
anonymized or are identifiable. This dichotomous representation
leaves out pseudonymized data. The participants were not
educated about the concepts of anonymization prior to answering
the questions potentially allowing ambiguity of the terminology
to cause a strong bias in the response behavior.

In addition, the sample is concentrated on people with diabetes
and those with other diagnosed health conditions, rather than
the general population. This is a strength as well as a weakness
as it focuses on a population with particular circumstances and
perspectives and one that may be more representative of a
post-COVID-19 population that has been sensitized to a
continuing health concern than the current population. This in
turn means that a comparison of these findings with a similar survey of the general population can shed light on the potential impact of awareness-raising policies.

Another strength of the study is the high number of respondents who completed the questionnaire. In total, 4764 people participated in the study. This provides a unique insight into the views of a population deemed as being at the highest risk of severe disease and mortality related to COVID-19 [46]. There was a skew in the representation of the demographic distribution of individuals in the nationwide population of people living with chronic health conditions, since White people were overrepresented in our sample (n=4434, 93.4%) but not overrepresented among those with diabetes more generally. While our survey focused on those with diabetes, the results provide novel insight into concepts crucial for societal trust in data use and sharing initiatives.

While the study design did not allow us to ascertain whether technology use itself was correlated with higher acceptance of data-sharing, such an analysis is possible and will be an important topic for future research.

The study’s findings suggest potential targets for further study and possible considerations for policy makers. There are two main implications: storing and processing data in pseudonymized form and emphasizing the use of synthetic data (generated from models estimated from real data but not involving any actual or identifiable human beings).

Understanding attitudes toward data sensitivities and trust can contribute to developing policies, improving transparency, and increasing the trust, speed, focus, and effectiveness of epidemic responses. Future practice should emphasize transparent data-sharing and privacy initiatives, while research should evaluate whether this does indeed lead to greater levels of trust and engagement. Encouraging ethical and relevant data-sharing can provide significant epidemic intelligence and support public health emergency relief operations [47].

Conclusions
Data sensitivity is highly contextual. More people are comfortable with sharing anonymized data than personally identifiable data. Willingness to share data also varied depending on the receiving body, highlighting trust as a key theme, who may have access to shared personal health data and how it may be used in the future. The nascent legal guidance in this area suggests the requirement for humanitarian guidelines for data responsibility during disaster relief operations such as pandemics, and the requirement to involve the public in their development.

Acknowledgments
The authors acknowledge and thank the Diabetes.co.uk Forum community, which supported this study.

Conflicts of Interest
CS and AP are founding employees of DDM, who operate the Diabetes.co.uk Forum community.

References

28. Gerber N, Gerber P, Volkamer M. Explaining the privacy paradox: A systematic review of literature investigating privacy


25. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES).


[doi: 10.1038/s42256-020-0194-1]


Abbreviations

NHS: National Health Service
Development of a Quality Management Model and Self-assessment Questionnaire for Hybrid Health Care: Concept Mapping Study

Rosian Tossaint-Schoenmakers1,2,3, MSc; Marise J Kasteleyn2,3, PhD; Anneloek Rauwerdink4, MSc, MD; Niels Chavannes2,3, MD, PhD; Sofie Willems2,3, MD, PhD; Esther P W A Talboom-Kamp2,3,5, MBA, MD, PhD

1Saltro Diagnostic Centre, Unilabs Netherlands, Utrecht, Netherlands
2National eHealth Living Lab, Leiden University Medical Centre, Leiden, Netherlands
3Department of Public Health and Primary Care, Leiden University Medical Centre, Leiden, Netherlands
4Department of Surgery, Gastroenterology and Metabolism, Amsterdam University Medical Centre, Amsterdam, Netherlands
5Unilabs Group, Geneve, Switzerland

Corresponding Author:
Rosian Tossaint-Schoenmakers, MSc
Saltro Diagnostic Centre, Unilabs Netherlands
Mississippi dreef 83
Utrecht, 3565 CE
Netherlands
Phone: 31 655895361
Email: rtossaint@saltro.nl

Abstract

Background: Working with eHealth requires health care organizations to make structural changes in the way they work. Organizational structure and process must be adjusted to provide high-quality care. This study is a follow-up study of a systematic literature review on optimally organizing hybrid health care (eHealth and face to face) using the Donabedian Structure-Process-Outcome (SPO) framework to translate the findings into a modus operandi for health care organizations.

Objective: This study aimed to develop an SPO-based quality assessment model for organizing hybrid health care using an accompanying self-assessment questionnaire. Health care organizations can use this model and a questionnaire to manage and improve their hybrid health care.

Methods: Concept mapping was used to enrich and validate evidence-based knowledge from a literature review using practice-based knowledge from experts. First, brainstorming was conducted. The participants listed all the factors that contributed to the effective organization of hybrid health care and the associated outcomes. Data from the brainstorming phase were combined with data from the literature study, and duplicates were removed. Next, the participants rated the factors on importance and measurability and grouped them into clusters. Finally, using multivariate statistical analysis (multidimensional scaling and hierarchical cluster analysis) and group interpretation, an SPO-based quality management model and an accompanying questionnaire were constructed.

Results: All participants (n=39) were familiar with eHealth and were health care professionals, managers, researchers, patients, or eHealth suppliers. The brainstorming and literature review resulted in a list of 314 factors. After removing the duplicates, 78 factors remained. Using multivariate statistical analyses and group interpretations, a quality management model and questionnaire incorporating 8 clusters and 33 factors were developed. The 8 clusters included the following: Vision, strategy, and organization; Quality information technology infrastructure and systems; Quality eHealth application; Providing support to health care professionals; Skills, knowledge, and attitude of health care professionals; Attentiveness to the patient; Patient outcomes; and Learning system. The SPO categories were positioned as overarching themes to emphasize the interrelations between the clusters. Finally, a proposal was made to use the self-assessment questionnaire in practice, allowing measurement of the quality of each factor.

Conclusions: The quality of hybrid care is determined by organizational, technological, process, and personal factors. The 33 most important factors were clustered in a quality management model and self-assessment questionnaire called the Hybrid Health Care Quality Assessment. The model visualizes the interrelations between the factors. Using a questionnaire, each factor can be
assessed to determine how effectively it is organized and developed over time. Health care organizations can use the Hybrid Health Care Quality Assessment to identify improvement opportunities for solid and sustainable hybrid health care.

**Introduction**

**Background**

In recent years, the use of eHealth has expanded, encouraged by the increasing pressure on health care [1,2] and growing interest in patient empowerment [3,4]. On the one hand, an aging population and an increase in chronic diseases are causing a higher and more complex demand for health care. In addition, the COVID-19 pandemic has accelerated pressure on health care [5-8]. Therefore, innovations such as eHealth are required to maintain accessibility and high quality of health care [9-12]. On the other hand, digital health technologies have significantly accelerated patients' involvement [13-16]. In line with these developments, health care organizations have intensively integrated eHealth into traditional face-to-face consultations [17]. The combination of eHealth and face-to-face consultations can be defined as hybrid health care [18,19]. A few examples of hybrid health care are telemonitoring systems for patients with chronic diseases [20,21], web-based video coaching [22,23], and direct web-based access to medical records of patients [24,25], all of which are integrated into traditional health care.

Although health care organizations are increasingly providing hybrid health care, integrating eHealth into the daily care process is challenging. Working with hybrid health care requires organizations to change the way they work. The roles of health care providers and patients are changing, and the available resources are used differently [4,22,26,27]. Organizational structure and work processes must be adapted to ensure high-quality hybrid care [28-31]. Several studies have examined ways to promote eHealth adoption, such as increasing the adaptability of the technology or stakeholders’ value [32,33]. However, it remains challenging to organize hybrid health care effectively and sustainably [17]. There is a need for further research on how hybrid health care can be improved to add value to patients and health care providers when they work with eHealth. Therefore, we recently performed a systematic literature review to optimally organize hybrid health care [17].

In the systematic literature review, the Donabedian Structure-Process-Outcome (SPO) framework was used to identify indicators related to the integration of eHealth into health care organizations [17,34-36] (Figure 1). According to Donabedian, health care quality is based on the aspects of these 3 categories and their relationships. The SPO framework and its categories are described in detail in a literature review [17].

In the literature review, we identified 111 potential indicators under the SPO categories that impact eHealth integration. The study demonstrated that 3 principles are important for successful integration. First, the patient’s role must be centrally placed in the organization of hybrid care. Second, technology must be well attuned to the organizational structure and daily care process. Third, the deployment of human resources must be aligned with desired results [17].

**Objectives**

To translate the findings from the literature study into a modus operandi for health care organizations, we aimed to develop a model that can help health care organizations organize hybrid health care and identify improvement opportunities for a solid and sustainable integration of eHealth. To achieve this aim, the objectives of the concept mapping study included the following: (1) enrich and validate evidence-based knowledge from the literature review with practice-based knowledge from experts and (2) develop an SPO-based model for organizing hybrid health care with an accompanying self-assessment questionnaire.

**Methods**

**Concept Mapping**

Concept mapping is a highly structured methodology for organizing ideas from different stakeholders and other data sources to produce a common framework for complex topics that can be used for evaluation or planning [37-40]. The method...
integrates qualitative data collection with quantitative analysis to construct an interpretable pictorial view of different ideas and concepts and how these are interrelated [41,42]. Concept mapping has been used worldwide, for a diverse range of health care projects and studies to develop conceptual frameworks, as well as health and eHealth evaluations [43-49].

In this study, the 6-step concept mapping approach of Trochim and McLinden [42] was followed [49] to develop a usable, tailored, SPO-based quality management model for hybrid health care and an accompanying questionnaire. The six steps of concept mapping are as follows: (1) preparation, (2) idea generation, (3) sorting and rating, (4) concept mapping analysis, (5) map interpretation, and (6) utilization. Each step involves different activities leading to an output, which serves as an input for the next step. The steps and activities are explained in Figure 2 and in the paragraphs below. All the steps were supported by the GroupWisdom webtool [41,42].

Figure 2. Concept mapping steps and study activities.

**Step 1: Preparation**

Concept mapping is most effective when multiple stakeholders participate in all the steps of the concept mapping process [50]. There is no strict limitation to the number of participants, ranging from small groups of 8 to 15 people to groups of hundreds of participants [50]. For this study, participants with eHealth experience, those employed by health care organizations, and patients with eHealth experience were recruited. The amount or kind of eHealth experience, health care setting, or disease was not relevant for inclusion. The goal was to create a diverse group in which different experiences, perceptions, and viewpoints complemented each other. We aimed to include a mix of health care professionals, patient experts (patients and caregivers), managers, directors, project leaders, researchers, and eHealth suppliers.

Potential participants were approached to attend both brainstorming in step 2 and sorting and rating in step 3. Participants were invited via the research team’s network, social media, and snowballing. Before agreeing to participate, participants received an information letter about the concept mapping method, the study’s purpose, and the SPO framework. None of the potential participants were familiar with our...
previous literature study results. A selected group was asked to participate in step 4 (concept mapping), step 5 (interpretation), and step 6 (utilization), which will be explained in the subsequent sections.

**Step 2: Idea Generation**

**Web-Based Brainstorming**

In step 2, data from the participants were collected and combined with data from the literature study. Idea generation with participants was organized by brainstorming. Brainstorming is the most common method used in concept mapping, and can be either group brainstorming or individual brainstorming [42]. In this study, web-based brainstorming was conducted by the participants. Participants received a link via email with instructions, giving them access to the web-based brainstorm program of the GroupWisdom webtool. Before starting the brainstorming session, informed consent was provided, and participant characteristics (age, eHealth experience, professional background, and work setting) were collected to generate general background information about the participants. When the brainstorming started session, the following instruction was presented: “Name all factors, which you believe contribute to effective organization of patient care with eHealth, and what the outcomes of this care should be. Keep the ‘Structure-Process-Outcome’ framework in mind.”

For 23 days, the participants could list as many factors they considered essential contributors to effective hybrid health care. Participants could see each other’s inputs and save their brainstorming results in the meantime. They received reminders after 10 and 15 days.

**Editing Brainstorming and Literature Study Data**

After closing the web-based brainstorming session, the brainstorming and literature study data were combined for sorting and rating. A manageable amount of data for sorting and rating is ideally ≤100 to prevent redundancy and a loss of participants’ motivation [51,52]. To generate a final set of up to 100 factors, duplicates and factors that did not match the brainstorming instructions were removed. For this purpose, each factor was assessed independently by the authors, RT-S and ET-K. The assessments were compared, and disagreements were resolved by discussion between RT-S and ET-K. Next, RT-S edited the remaining factors for grammar and spelling. Authors, MK and AR reviewed the editing process to check whether they would conclude the same selection and wording and made recommendations where appropriate. Finally, the set was entered into the GroupWisdom webtool, serving as an input for the sorting and rating activities.

**Step 3: Sorting and Rating**

At the beginning of step 3, the participants received instructions for the sorting and rating tasks. For the sorting task, the participants were asked to cluster the factors into self-created clusters and assign names to the clusters. The participants were instructed to keep the Donabedian SPO categories in mind while sorting each factor into self-created clusters. For the rating task, each participant was asked to rate each factor by relevancy on a 5-point Likert scale, ranging from 1 (not important at all or not feasible to measure) to 5 (very important or very feasible to measure) by answering the questions, “How important is this factor for effective patient care with eHealth?” and “How feasible to measure is this factor?”

The participants had the opportunity to sort and rate over 3 weeks. They could save their activities and return later and received reminders after 10 and 15 days. The sorting data were approved for concept mapping analysis for participants who completed 75% of the sorting activity and created at least three clusters [41]. The rating data were included when the participant rated at least one factor.

**Step 4: Concept Mapping Analysis**

Concept mapping analysis consisted of four main activities: (1) generating a point map with the sorting data, (2) grouping factors into clusters using hierarchical cluster analysis, (3) selecting a concept map from the hierarchical cluster analysis, and (4) computing average ratings for each factor and cluster of the selected concept map [50]. All computations were based on the concept mapping approach of Kane et al [53,54] and conducted using the GroupWisdom webtool.

**Generating a Point Map With the Sorting Data**

Data from the rating step were analyzed to create a point map [45,53,55,56]. A point map is a 2-dimensional point map, in which each point represents a factor [53]. The point map visually displayed the locations of all factors. Factors closer to each other on the point map were sorted together more frequently by the participants, whereas more distant factors on the map were sorted together less frequently [42,50,53]. The point map was constructed using a similarity matrix and multidimensional scaling algorithm. First, the similarity matrix indicated the number of times various factors were grouped together. Next, a multidimensional scaling algorithm plotted factors as points on a point map [42,54,55]. Subsequently, a stress value (0-1) was calculated, indicating the degree to which the distances on the point map fit the original similarity matrix [38,54]. The better the fit, the lower is the stress value.

**Grouping Factors Into Clusters With Hierarchical Cluster Analysis**

The point map provided the input for the hierarchical cluster analysis. The hierarchical cluster analysis grouped factors into clusters [44] using Ward algorithm [57]. The algorithm proposed several concept map solutions, where 2 clusters were merged at each following the proposed solution.

**Selecting a Concept Map**

From the proposed concept map solutions, a concept map that made sense for conceptualization was selected. There is no single correct number of clusters or mathematical decision criterion for selecting a concept map solution [38,56]. This study selected the number of clusters for the concept map by determining the range of the highest and lowest number of clusters. The range was the average number of clusters made by the participant and its SD.

Subsequently, the cluster solutions in this range were reviewed to select the cluster level by following the cluster tree in the Methods section of the studies by Trochin [53] and Kane et al.
Finally, in a meeting, 2 authors (RT-S and ET-K) and 2 participants reviewed the merging of clusters, beginning with the highest number of clusters and moving to the lowest. The 2 study participants were asked to join this meeting because of their extensive experience with eHealth, daily care processes, research, operational management, and concept mapping.

After establishing the number of clusters in the concept map, each factor was reviewed for compatibility with the cluster and to determine whether it was appropriate to move the factor to a different cluster. A cluster and its content were appropriate for inclusion when they were considered essential and usable for the quality management model [53].

In addition, each cluster received a name and description based on the cluster names that emerged from the sorting activity.

**Computing Mean Ratings for Each Cluster and Factor of the Selected Concept Map**

After the cluster map was selected, the relationships between ratings were computed using pattern-match and Go-zones [42]. Pattern-match and its Pearson product-moment (r value) were calculated to compare how the clusters of the selected concept map were rated on importance and measurability. The pattern-match visualized the mean ratings of each cluster in a ladder graph, connecting lines between the mean ratings on importance and measurable of each cluster [50,57]. The r value represented the correlation strength between the 2 mean ratings of all clusters [50,57].

Finally, multiple Go-zones were computed: a Go-zone of the total point map and Go-zones per cluster of the selected concept map. Go-zone is a 4-quadrant graph with an x-y graph [50], visualizing the mean ranking results of each factor on the questions “How important is this factor” and “How feasible to measure is this factor.” The minimum and maximum values for each axis were the minimum and maximum average Likert scores, respectively. The upper-right quadrant is called the Go-zone because it shows factors rated above the mean for both importance and measurability [42,58]. The pattern-match and Go-zone showed how important and measurable each cluster and its factors were rated for quality assessment by the individual participants during the step, sorting and rating.

The selected concept map, with its calculation of importance and measurability for each cluster and factor, formed the basis of interpretation in the next step [53].

**Step 5: Interpretation of the Concept Map**

The selected concept map, with its pattern-match and Go-zones, was discussed with an advisory board. On the basis of the pattern-match and Go-zones, the advisory board decided which clusters and factors should be included in the quality management model and the accompanying questionnaire. The advisory board consisted of 4 study participants from the brainstorming and sorting step, of whom, 2 also participated in step 4, concept mapping analysis. The advisors were chosen because they could be future model users. In addition, all had extensive experience with eHealth, health care business, and as health care professionals (general practitioners, nurses, anesthetists, and clinical psychologists) in different health care settings.

The advisors voted individually on which clusters and factors of the selected concept map should be included in the quality management model and questionnaire to ensure usability. Using a web-based survey, the following questions were asked: “Which cluster should be included in the quality management model based on the mean cluster rating scores of the pattern matches? Please, specify your choice.” and “On which factors should the questionnaire give focus? Guide your choice by the Go-zones of each cluster and the Go-zone of the total point map. Please specify your choice.” The advisors could not see each other’s votes. By 75% (3/4) agreement or more, the concerned clusters and factors were operationalized in the quality assessment model and questionnaire. Where there was less agreement, the advisors viewed all responses, including the comments, and were asked to vote again. This process was repeated until a 75% consensus was reached. The web-based survey results were used as inputs to develop the quality management model and its questionnaire.

**Step 6: Utilization**

**Quality Management Model**

The remaining clusters and their positions in the selected concept map provided the blueprint for the quality management model. First, the excluded clusters and factors were removed from the concept map. Second, the concept map with the remaining clusters was used to produce a logic model. A logic model is a framework that visualizes the interrelations between the clusters in graphic form and is therefore valuable for quality evaluation [59]. The SPO framework [34,35] was used to identify logical interrelationships between the clusters. Accordingly, noticeable SPO connections between the clusters were drawn on the map by RT-S. A simplified version of the logic model was designed for clarity and readability. Authors SW, ET-K, and RT-S discussed the design of the quality management model to ensure the usability and clarity of the model.

**Self-assessment Questionnaire**

The questionnaire was drafted by RT-S with the remaining factors, taking the advisors’ comments into account. The questionnaire should give care organizations insight into the quality of hybrid care and how quality develops over time. On the one hand, the questionnaire must be easy to use and uniformly independent of the type of health care organization, type of eHealth, and disease. On the other hand, the questionnaire results must provide specific guidance to improve the quality of specific clusters and factors.

The concept model and questionnaire were submitted to the advisors for peer review of usability and clarity. Their comments were processed by RT-S, resulting in an improved draft. Finally, ET-K and SW peer reviewed the last draft to ensure that the representatives’ comments were implemented entirely in the quality management model and the related questionnaire.

**Ethics Approval**

Approval by an ethics committee was not needed because no intervention or trial has occurred in the sense that the research...
participants were subjected to actions or had modes of behavior imposed on them [60].

Results

Participant Characteristics (Step 1)

A total of 39 people participated in this study. The participants had a mean age of 45.2 (SD 11.1) years and were mainly working at the family medicine clinic (12/39, 31%) or hospital (10/39, 26%) within a management function (16/39, 41%) or as a health care professional (14/39, 36%). A total of 59% (23/39) of the participants estimated their eHealth experience to be extensive. The 3 most commonly used eHealth tools were apps (37/147, 25.2% participants), web portals (35/147, 23.8% participants), and video communication (34/147, 23.1% participants). An overview of the participants’ characteristics is shown in Table 1.

Of the 39 participants, 38 (97%) completed the brainstorming sessions. In all, 18% (7/38) of the participants dropped out after the brainstorming session, and a new participant joined the sorting and rating phase. In total, 79% (31/39) of the participants completed the sorting and rating phase (Figure 3).

Table 1. Participant characteristics (N=39).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>45.2 (11.1)</td>
</tr>
<tr>
<td>Main work setting, n (%)</td>
<td></td>
</tr>
<tr>
<td>Family medicine</td>
<td>12 (31)</td>
</tr>
<tr>
<td>Hospital</td>
<td>10 (26)</td>
</tr>
<tr>
<td>Mental health clinic</td>
<td>5 (13)</td>
</tr>
<tr>
<td>Nursing and residential care</td>
<td>5 (13)</td>
</tr>
<tr>
<td>eHealth supplier</td>
<td>4 (10)</td>
</tr>
<tr>
<td>Research institute</td>
<td>2 (5)</td>
</tr>
<tr>
<td>Patient experts (self-employed)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Main profession, n (%)&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Manager, director, or project leader</td>
<td>16 (41)</td>
</tr>
<tr>
<td>Health care professional (eg, physician, nurse, therapist, or psychologist)</td>
<td>14 (36)</td>
</tr>
<tr>
<td>Patient expert (eg, patient or caregiver)</td>
<td>5 (13)</td>
</tr>
<tr>
<td>Researcher</td>
<td>3 (8)</td>
</tr>
<tr>
<td>Unknown</td>
<td>1 (3)</td>
</tr>
<tr>
<td>eHealth technology experience, n (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Apps</td>
<td>37 (25.2)</td>
</tr>
<tr>
<td>Web portals (eg, electronic health records or personal care records)</td>
<td>35 (23.8)</td>
</tr>
<tr>
<td>Video communication</td>
<td>34 (23.1)</td>
</tr>
<tr>
<td>Sensors and wearables</td>
<td>23 (15.6)</td>
</tr>
<tr>
<td>Artificial intelligence</td>
<td>13 (8.8)</td>
</tr>
<tr>
<td>Domotica and robotica</td>
<td>10 (6.8)</td>
</tr>
<tr>
<td>Estimated level of experience with eHealth, n (%)</td>
<td></td>
</tr>
<tr>
<td>Extensive experience</td>
<td>23 (59)</td>
</tr>
<tr>
<td>Moderated experience</td>
<td>15 (38)</td>
</tr>
<tr>
<td>Limited experience</td>
<td>1 (3)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Many participants had dual roles, from which they were asked to choose one role.

<sup>b</sup>Participants could select multiple answers.
Idea Generation (Step 2)

Brainstorming during idea generation resulted in a list of 203 factors. A total of 111 potential indicators were extracted from the literature study [17]. Both lists were aggregated, resulting in a list of 314 factors. Editing of the data led to a final list of 78 factors. These 78 factors served as inputs for the sorting and rating activity. The list of 78 factors is provided in Multimedia Appendix 1.

Sorting and Rating (Step 3)

The rating data of the 32 participants were included in this study. All factors received mean rating scores of >3.1, for both importance and measurability. The mean ratings on the questions, “How important is this factor for successful integration of eHealth?” and “How feasible to measure is this factor” are described in Multimedia Appendix 1.

Selecting the Concept Map

Concept map solutions ranging from 11-cluster to 3-cluster options were reviewed (mean 7, SD 3.5). The 9-cluster concept map was selected to make the most sense of conceptualization. A few factors (n=14) were unanimously replaced, leading to the concept map shown in Figure 5. Replaced factors and their reasons are presented in Multimedia Appendix 2. The 9 clusters were labeled and received a short description, as described in Table 2. The number of points corresponds to the number of factors presented in Multimedia Appendix 1. The clusters represent how the participants sorted the factors into self-created clusters using the proposed cluster labels.
Table 2. Clusters labels and descriptions.

<table>
<thead>
<tr>
<th>Cluster number</th>
<th>Cluster label</th>
<th>Description</th>
<th>Included factors, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Quality information technology infrastructure and systems</td>
<td>Conditions concerning technology, information technology systems, and data.</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>Quality eHealth application</td>
<td>Conditions concerning the eHealth application.</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>Learning system: evaluation and improvement</td>
<td>Evaluation and realignment with stakeholders and the patient care objectives for a continuous development.</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>Vision, strategy, and organization</td>
<td>Responsibilities of the health care organization concerning vision, strategy, policy, leadership, funding, and work process designs.</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>Providing support to health care professionals</td>
<td>Conditions arranged by the health care organization to encourage the use of eHealth among its health care professionals.</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>Skills, knowledge, and attitude of health care professionals</td>
<td>Health care professionals’ ability to provide hybrid care.</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>Attentiveness to the patient</td>
<td>Organize the daily care process in line with the patient’s needs, demand for care, and its capacity.</td>
<td>13</td>
</tr>
<tr>
<td>8</td>
<td>Organization outcomes</td>
<td>Outcomes for the health care organization; for example, quality health care provision and health care logistics.</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>End results for the patient</td>
<td>Outcomes for the patients; for example, health, added value, satisfaction, ownership, and convenience.</td>
<td>10</td>
</tr>
</tbody>
</table>

Number corresponds with the number of the concerning cluster in Figure 5.

Mean Ratings for Each Cluster and Factor of the Selected Concept Map

The pattern-match showed that all clusters had a mean score between 3.75 and 4.27 on the importance and a mean score between 3.79 and 4.10 on measurability (Figure 6). The cluster with the highest mean score on importance was Attentiveness to the patient (mean 4.27, SD 0.27), and the cluster with the highest mean score on measurability was End results for the patients (mean 4.10, SD 0.17). On the contrary, the cluster with the lowest mean score on importance was Organization outcomes (mean 3.75, SD 0.36), whereas the cluster Quality eHealth application (mean 3.79, SD 0.45) had the lowest mean score on measurability. The r value was 0.63, indicating a predictable alignment between the rating of importance and the rating of measurability. The mean ratings of the factors and Go-zones per cluster are included in Multimedia Appendix 3.
Interpretation of the Concept Map (Step 5)
The pattern-match and Go-zones were input to determine which clusters and factors of the selected concept map should be included in the quality management model and questionnaire. Decisions were made in 2 voting rounds. Of the 9 clusters, the cluster Organization outcomes was not included in the quality management model, based on the voting (3/4, 75%) of the advisors had doubts about including the cluster in the model) and after discussion with the research team. The factors included in the questionnaire concerned those placed in the Go-zone of the total point map or the Go-zone of the clusters. As a result, 8 clusters remained in the model and 33 factors in the questionnaire remained as a manageable utility for quality assessment (Textbox 1). Multimedia Appendix 3 presents the responses and comments of the advisory board during the voting rounds.
**Textbox 1. The included clusters and factors.**

**Quality Information technology infrastructure and systems (1)**
- Information technology architecture available within the health care organization (1).
- Back-up scenario during technical problems (12).

**Quality eHealth application (2)**
- The eHealth application is user-friendly (35).

**Learning system: evaluation and improvement (3)**
- Cocreation: eHealth is developed, implemented and redeveloped with different stakeholders (8).
- Monitoring and evaluation of service and treatment results (58).

**Vision, strategy, and organization (4)**
- Support the implementation and development of eHealth in the organization with good project management (4).
- Mobilizing funding for working with eHealth (16).
- Clear internal policies regarding the use of eHealth (18).
- Vision supported by the line, “Why are we doing this?” (21).
- Care delivery with eHealth complies with laws and regulations (41).
- Financial reimbursements for eHealth deployment (42).
- Redesign the current work process and review what contributes to the desired care outcomes (47).

**Providing support toward health care professionals (5)**
- Health care professionals have easy access to information technology resources; for example, device, internet, screen, or headset (2).
- Embedding eHealth in the daily practice of health care professionals (11).
- Training and supervision for health care professionals (15).
- Help desk for health care professionals (17).
- Information on the treatment with eHealth is clear and accessible to the health care professional (19).

**Skills, knowledge, and attitude of health care professionals (6)**
- Good balance between face to face and eHealth for the health care professional (46).
- The health care professional has confidence in the eHealth application (70).
- The health care professional is satisfied with working with eHealth (74).

**Attentiveness to the patient (7)**
- Clear communication to the patient about how care is offered (10).
- Personalized care, considering patient needs with regard to (deployment of) eHealth (13).
- The patient has easy access to the necessary information technology resources; for example, device, Internet, and so on (30).
- Patients receive practical support in using the eHealth application; for example, a help desk (49).
- The patient has confidence in the eHealth application (67).
- The patient has the flexibility to use eHealth wherever and whenever it is convenient (72).

**End results for the patient (9)**
- The patient can integrate the use of eHealth in their daily life (33).
- Treatment with eHealth has a positive influence on the patient’s health (64).
- Treatment with eHealth contributes to the patient’s self-reliance (65).
- The patient is satisfied (68).
- The patient has easy access to care (71).
- eHealth provides logistical convenience for the patient (73).
eHealth has added value for the patient (75).

Utilization (Step 6)

Utilization Model

The clusters and factors excluded from the voting rounds were removed from the selected concept map. The remaining clusters (n=8) and their factors (n=33) led to nonoverlapping clusters on the concept map. Above the clusters, the SPO categories were positioned as overarching themes to emphasize the interrelations between the clusters. In addition, a complex cluster map can be simplified into a logic model. Figures 7A-C show the simplification of the model.

The overarching categories, structure, process and outcomes and the clusters’ interconnections refer to the Donabedian SPO framework [34,35]. The cluster Learning system is visualized in the arrows with the dashed line. The numbers inside the clusters represent the number of factors included.

Figure 7. Simplification of the model. (A) Removing the excluded cluster and factors from the selected concept map and adding the overarching categories’ structure, process, and outcome. (B) Drawing a logic interrelationship with structure, process, and outcome categories. (C) Simplification into a quality management model. IT: information technology.

Utilization Questionnaire

The remaining 33 factors were included in the questionnaire, where each factor can be measured on how effectively it is organized and developed over time. The advisory board noted that measuring the quality progress of hybrid health care is very important, in addition to learning and continuous improvement with stakeholders. Subsequently, the idea was to enrich the questionnaire with a quality progress tracker based on the plan-do-check-act (PDCA) cycles of Deming [61]. Incorporating the PDCA cycle makes it possible to assess the quality easily.
and uniformly with tailored feedback for health care organizations. PDCA is a well-known cycle method for continuous improvement and quality measurement [61]. The PDCA cycles assess each factor’s quality by measuring the extent to which The objective is tangible? (plan), The plan is implemented? (do), To what extent is the plan realized? (check), andProviding feedback on the quality of the execution to make improvements (act) [61]. Each factor can be monitored on the quality level of the PDCA cycles using a Likert score (0-10). A score of 0 means there is no plan to improve the concerning factor, and a score of 10 means continue improvement with stakeholders. The Likert scoring is based on the PDCA cycles and the 2 factors of the cluster Learning system, which include the following: (1) Co-creation: eHealth is being developed and implemented with various stakeholders and (2) Monitoring and evaluation of service- and treatment outcomes. Using the PDCA cycles in combination with a Likert score provides a health care organization insight into improvement possibilities for each factor or cluster.

Finally, the model and questionnaire obtained a more convenient workname Hybrid Health Care Quality Assessment (HHQA). The HHQA model and questionnaire with suggestions on how to use it are explained in Multimedia Appendix 4.

Discussion

Principal Findings

In this concept mapping study, we aimed to develop an SPO-based model and an accompanying self-assessment questionnaire for hybrid health care. By combining practice-based knowledge from eHealth users with an evidence-based literature review, we found that organizational, technological, and process and personal factors affect the quality of hybrid health care. Health care organizations must understand that these factors play a role in organizing hybrid health care and should be familiar with ways to improve them. The authors developed the HHQA, which can be used to systematically assess and improve the quality of hybrid health care.

The HHQA model includes 8 clusters. Cluster 1 (Vision, strategy, and organization) includes the responsibilities of the management to set the vision, strategy, policy, leadership, finance, and project management. Cluster 2 (Quality information technology infrastructure and systems) focuses on information technology infrastructure and back-up scenarios by information technology issues. Cluster 3 (Quality eHealth application) concerns the user-friendliness of the digital health application itself. Cluster 4 (Providing support toward care professional) and cluster 5 (Skills, knowledge, and attitude of health care professionals) include factors concerning health care providers. Cluster 4 focuses on factors that should be arranged for the individual health care professional by the care organization, and cluster 5 includes the responsibilities of the professional. The patient is central in cluster 6 (Attentiveness to the patient). This cluster contains the measurement of factors that allow patients to increase their self-management and consider the individual patient’s needs. Patient centeredness is also reflected in cluster 7 (Patient outcomes), including factors such as patient’s health outcomes, added value, satisfaction, ownership, and convenience. Finally, cluster 8 (Learning system), forms the relationship between the continued development of hybrid health care with stakeholders and health care provision objectives. The factors in cluster 8 provide insight into where alignment can be improved with other organizational criteria and actions, such as cost-benefit or capacity management.

The interdependencies of the clusters are logically expressed in the HHQA model because of the overarching categories of the Donabedian SPO framework. Moreover, according to eHealth users, clusters consist of the most important factors for the quality of hybrid health care. Using the questionnaire, each factor (33 in total) was measured to determine how effectively it was organized and developed over time. Subsequently, the main results of the questionnaire were shown at the cluster level. It was possible to zoom in on the relevant factors for each cluster.

Comparison With Literature

In our previous literature review [17], we concluded that the capabilities of patients, health care professionals, and technology play a crucial role in the quality of hybrid health care. We also concluded that offering hybrid health care requires adjusting the daily care process and appropriate process monitoring. The conclusions from the literature review are reflected in the HHQA clusters, namely, the patient’s role is visible in the clusters Attentiveness to the patient and Patient outcomes; the health care professional’s role is central in the clusters Providing support toward health care professionals and Skills, knowledge, and attitude of professionals; and technology is covered in the clusters Quality information technology infrastructure and systems and Quality eHealth application. The adjustment of the daily care processes is elaborated in the cluster Vision, strategy, and organization. Finally, monitoring is embedded in the cluster Learning system and the PDCA-progress tracker.

The 8 clusters of the HHQA model fit the 3 overarching categories of the Donabedian SPO framework. According to Donabedian [34], health care quality is based on aspects of these 3 categories and their relationships. The interaction between the categories can be bidirectional and is an “unbroken chain of antecedents, followed by intermediate ends, which are themselves the means to still further ends” [35]. Our research translated the complex interaction between the categories, structure, process, and outcome into user language.

The HHQA connects essential contributions to the quality of hybrid health care using a progress tracker. The relationship between quality contributors and continuous improvement also appears in the European Foundation for Quality Management Model (EFQM) [62,63]; nonadoption, abandonment, scale-up, spread, sustainability (NASSS) [32]; and the Consolidated Framework for Implementation Research (CFIR) [64,65]. All models approach the organizational structure, process, and outcomes with continuous improvement in a structured manner, but with different focus areas. For example, the EFQM is not specified for health care, in contrast to the NASSS and CFIR. The NASSS focuses on the adoption of technology and reduces implementation complexity, whereas the CFIR emphasizes on implementation in general. However, none of them have been
specified for quality assessment and improvement of hybrid health care.

Nevertheless, it is interesting to conduct a detailed examination of the assessment questionnaires of the EFQM and NASSS. The EFQM deployed the Results-Approach-Deployed-Assessment-Refinement (RADAR) method [66,67], a questionnaire to assess the quality improvement at each EFQM criteria, which incorporates the continued improvement circle. The assessment using the RADAR method is similar to the PDCA cycle in our questionnaire, as both monitor continuous quality improvement by completing the cycle plan-executing-monitoring and refining. However, the RADAR, similar to the EFQM model, is not specified for hybrid health care. In addition, the NASSS comes with a questionnaire to monitor the complexity of technology implementation in health care [68], but the focus is on project management instead of the hybrid health care process itself. Furthermore, there are other questionnaires measuring the quality of eHealth [69-72] or the quality of health care [73,74]. However, these questionnaires are concerned with the quality assessment of eHealth nationwide [68,70], the quality of a specific digital health application [70,72], or measuring the quality of a specific disease pathway [73,74]. To the best of our knowledge, HHQA is the first questionnaire measuring the quality of hybrid health care at an organizational level, taking the role of the patient, health care professionals, and technology into account, accompanied by an improvement progress tracker. Therefore, the authors recommend using the HHQA to measure and improve the quality of hybrid health care.

Strengths and Limitations

This study has several strengths. First, the HHQA was developed in cocreation with stakeholders who are direct users of eHealth. Therefore, the HHQA content was drawn from inside the health care system itself and not conceived or imposed outside the health care organizations. Second, stakeholders choose the included clusters and factors. The researcher only played a facilitating role. Consequently, the clusters and factors accurately reflect stakeholders' views and values, expressed in their own words and visual representations. Third, the stakeholder group was diverse and consisted of representatives of health care professionals, patients, managers, researchers, and eHealth designers. Nevertheless, the stress value of the point map shows that the stakeholders' outcomes are highly compatible. Therefore, the study results are likely to be generalizable to everyday practices. Fourth, the model and questionnaire were developed by combining scientific and practice-based knowledge. Together, these strengths result in important factors for effective hybrid health care covering different users' needs and organization requirements.

Our study had some limitations. First, the questionnaire had not yet been tested in health care organizations. This will be conducted in a follow-up study. Although eHealth users from different health care organizations have reviewed the model and questionnaire, the model and questionnaire may still be too abstract for daily practice, as is often the case in scientific research [75-77]. A follow-up study could provide concrete recommendations on how to use the HHQA. Second, it is conceivable that other factors and clusters could be included in other participants and health care environments. We attempted to overcome this problem by creating diverse groups of participants with different backgrounds, various eHealth experiences, and different kinds of health care settings. In addition, combining idea generation through brainstorming with results from a systematic literature review reduces the risk of bias. Third, based on the analysis of the concept mapping phase, 14 factors were moved to other clusters. However, some of these factors were moved far across the map, which was not entirely in line with the spirit of group concept mapping. Nevertheless, we deemed it necessary to move these factors for substantive reasons. Fourth, the advisory group consisted of 4 participants. We wanted to avoid overquestioning the participants and, therefore, deliberately selected a group of delegates who reflected on the diversity among the participants and who also had experience with quality management and concept mapping. Combined with in-depth preparation and discussion among the research groups, this appeared to be the most feasible solution.

Finally, it is worth pointing out that the HHQA gives a first general impression of improvement, as there is much to be gained in taking the role of the patient, health care professionals, and used technology into account [17]. Furthermore, the authors will continue with follow-up research and warm-heartedly welcome repetition of the study to improve the HHQA, taking into account the different users and health care environments.

Conclusions

This study developed a quality management model and an accompanying self-assessment questionnaire tailored for hybrid health care, the HHQA. A quality model for hybrid care is indispensable for effectively integrating eHealth into regular care and delivering high-quality health care. The HHQA covers all relevant aspects for the assessment and sustainable improvement of hybrid health care and the interrelations of eHealth with organizational, technical, and human factors. The next step is to validate and apply the HHQA model and questionnaire in practice.

Acknowledgments

The authors thank the experts who participated in the brainstorming and sorting phase, Arjen Huizinga (MiGuide), Bart van Pinxteren (eHuisartsenkompas; Huisartsen Oog in Al; Huisartsen Utrecht Stad; Nederlands Huisartsen Genootschap; Spindok); Bart Timmers (Groepspraktijk Huisartsen Bergh), Beverly Rose (Vereniging van Ervaringsdeskundigen), Corine van Barneveld (Saltro), Caroline Meijer (Leiden University Medical Centre), Erwin van Boxtel (Thebe), Folkert de Winter (Sint Antonius Ziekenhuis), Geert-Jan van Hal (self-employed), Hans in het Veen (Franciscus Gasthuis and Vlietland), Heleen Krabben (Saltro), Caroline Meijer (Leiden University Medical Centre), Irvin Talboom (Huisartsen Zorggroep Breda), Jan Frans Mutsaerts (Het Huisartsenteam; Familiedokters), Jeanette Ploeger (Minddistrict), Joris Arts (DiSofa; Geestelijke GezondheidsZorg Noord Holland Noord), Joyce

https://formative.jmir.org/2022/7/e38683 JMIR Form Res 2022 | vol. 6 | iss. 7 | e38683 | p.273 (page number not for citation purposes)
Bierbooms (Geestelijke GezondheidsZorg Eindhoven), Judie Knol (de Neckar), Kim Brons (Leiden University Medical Centre; Process in progress), Leonoor van Dam van Isselt (Leiden University Medical Centre, Pieter van Foreest; Vilente), Maarten Ellenbroek (Topaz), Marijke de Vries (Leiden University Medical Centre), Marjolein Eldenhorst (Leiden University Medical Centre), Marijn Peters (Insight4u), Mieke Klerkx (self-employed), Lya van der Veen, Pascale Schure (Huisartsen Linschoten; Unilabs Group), Paul Haarkamp (Carinova), Ryan Esser (Incluzio), Stephanie Wouthuis (Gezondheidscentrum De Boog; Huisartsenpraktijk Nijdam; Saltro), Stijn de Ruijter (doccs huisartsenpraktijk), Wim Green (Leiden University Medical Centre).

The authors extend special thanks to Cynthia Hallensleben (Gezondheidscentrum Spoorlaan; Leiden University Medical Centre), Fred de Boer (Leiden University Medical Centre), Nathan Bachran (Geestelijke GezondheidsZorg Oost Brabant), and Tobias Bonten (Huisartsenpraktijk L Broek; Leiden University Medical Centre) for their advice on the development of the model and questionnaire, to Marc Smelik (IE Business School) for reviewing the manuscript textually, and to Léon Tossaint (former chief executive officer of the European Foundation for Quality Management Model) for his explanation of the Results-Approach-Deployed-Assessment-Refinement method of the European Foundation for Quality Management Model.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Mean (SD) rating scores of clusters and factors.

Multimedia Appendix 2
Relocation factors and their reasons.

Multimedia Appendix 3
Results voting “which clusters and factors to include” and given comments.

Multimedia Appendix 4
Suggestion utilization Hybrid Health Care Quality Assessment questionnaire.

References


41. Groupwisdom. URL: https://groupwisdom.com [accessed 2021-03-08]

42. Trochim WM, McLinden D. Introduction to a special issue on concept mapping. Eval Program Plann 2017 Feb;60:166-175. [doi: 10.1016/j.evalprogplan.2016.10.006] [Medline: 27780609]


63. European Foundation for Quality Management. URL: https://www.efqm.org/ [accessed 2022-03-04]


Abbreviations

CFIR: Consolidated Framework for Implementation Research
EFQM: European Foundation for Quality Management Model
Assessing the Views and Needs of People at High Risk of Gestational Diabetes Mellitus for the Development of Mobile Health Apps: Descriptive Qualitative Study

Beibei Duan1*, BSc; Zhe Liu1*, BSc; Weiwei Liu1, MD; Baohua Gou2, BSc

1School of Nursing, Capital Medical University, Beijing, China
2Beijing Youyi Hospital, Capital Medical University, Beijing, China
*these authors contributed equally

Corresponding Author:
Weiwei Liu, MD
School of Nursing
Capital Medical University
No. 10, Xi Tou tiao, You'an Men Wai, Feng tai District
Beijing, 100069
China
Phone: 86 13581694127
Email: lwwhlxy@ccmu.edu.cn

Abstract

Background: Early prevention of gestational diabetes mellitus (GDM) can reduce the incidence of not only GDM, but also adverse perinatal pregnancy outcomes. Moreover, it is of great significance to prevent or reduce the occurrence of type 2 diabetes. Mobile health (mHealth) apps can help pregnant women effectively prevent GDM by providing risk prediction, lifestyle support, peer support, professional support, and other functions. Before designing mHealth apps, developers must understand the views and needs of pregnant women, and closely combine users’ needs to develop app functions, in order to better improve user experience and increase the usage rate of these apps in the future.

Objective: The objective of this study was to understand the views of the high-risk population of gestational diabetes mellitus on the development of mobile health apps and the demand for app functions, so as to provide a basis for the development of gestational diabetes mellitus prevention apps.

Methods: Fifteen pregnant women with at least one risk factor for gestational diabetes were recruited from July to September 2021, and were interviewed via a semistructured interview using the purpose sampling method. The transcribed data were analyzed by the traditional content analysis method, and themes were extracted.

Results: Respondents wanted to develop user-friendly and fully functional mobile apps for the prevention of gestational diabetes mellitus. Pregnant women’s requirements for app function development include: personalized customization, accurate information support, interactive design, practical tool support, visual presentation, convenient professional support, peer support, reasonable reminder function, appropriate maternal and infant auxiliary function, and differentiated incentive function. These function settings can encourage pregnant women to improve or maintain healthy living habits during their use of the app.

Conclusions: This study discusses the functional requirements of target users for gestational diabetes mellitus prevention apps, which can provide reference for the development of future applications.

(JMIR Form Res 2022;6(7):e36392) doi:10.2196/36392

KEYWORDS
gestational diabetes mellitus; high-risk groups; mobile health; mHealth; applications; user-centered design; qualitative research

Introduction

Gestational diabetes mellitus (GDM) refers to the first occurrence of abnormal glucose tolerance of varying degrees during pregnancy, which is one of the most common complications of pregnancy [1]. Due to differences in diagnostic criteria and race, the global incidence of GDM is between 1% and 25% [2]. A previous study has shown that the incidence of GDM in China is 14.8%, and with the adjustment of the fertility
policy and the increase in elderly people and obese pregnant women, the incidence of GDM shows a trend of increasing every year [3,4]. GDM brings serious threats and challenges to both maternal and infant health. GDM can increase the risks of pregnancy hypertension, abortion, polyhydramnios, premature delivery, dystocia, and cesarean section [4-6]. Women who have a history of GDM have a 50%-73% risk of recurrence of GDM when they get pregnant again and a 10-fold higher risk of developing type 2 diabetes in 5 to 10 years after delivery, and the risk of cardiovascular disease is also higher in these women than in normal women [7]. Compared with the offspring of healthy women, the offspring of GDM patients have higher risks of fetal malformation, macrosomia, large for gestational age, hyperinsulinemia, neonatal hypoglycemia, pathological jaundice, and respiratory distress syndrome [8,9]. In addition, these children have increased risks of obesity, abnormal glucose tolerance, and type 2 diabetes when they become adults, bringing heavy economic burden to countries around the world [10]. Therefore, early preventive measures for pregnant women at high risk of GDM are of great significance in reducing the incidence of GDM, reducing adverse perinatal pregnancy outcomes, and preventing or reducing the occurrence of type 2 diabetes.

Mobile health (mHealth) refers to medical treatment and health management through mobile devices, such as mobile phones, patient health data monitoring devices, palm computers, and other wireless devices [11]. Mobile information technology can bridge the communication between medical service providers and users, helping doctors understand the health status of users and providing clinical decision support, so as to achieve remote diagnosis and treatment. At the same time, users’ needs for self-management support can be met to improve their compliance and self-management behaviors [12-14]. In addition, mobile information technology can greatly save users’ time and transportation costs, relieve the burden of medical treatment faced by hospitals and community health service institutions, and make the limited medical service resources the most effective. Therefore, mHealth has significant potential in the medical and health fields [15,16].

As one of the main forms of mHealth management, mHealth apps refer to health service platforms that use smartphones, tablets, and other mobile devices as terminals and rely on mobile internet technology to provide services for patients and medical staff [17]. With the popularity of mobile devices, more than 500 million smartphone users worldwide use mHealth apps for health management [18]. Due to their unique biofeedback function, mHealth apps can carry out real-time health assessment and provide feedback on patients’ health status, which can be considered beneficial for medical staff to implement personalized and precise health management for patients, and can have good application prospects [19]. At present, mHealth apps have been widely used in the prevention and health management of diabetes patients and have achieved satisfactory results [20]. However, currently, apps are mostly developed for commercial needs and purposes, the needs of patients are seldom taken into account during development, and few patients are invited to participate in the design process of apps [21-23]. Studies have shown that lack of interest in apps, fees, and fear of personal information disclosure are the main reasons that hinder users from downloading and sticking with apps [24]. Therefore, when developing mHealth apps, it is necessary to take users as the center, pay attention to users’ preferences and needs for using apps, and design apps that can meet users’ needs, so as to improve users’ experiences, increase users’ stickiness in using apps, and improve patients’ compliance with self-management.

User-centered design (UCD), which is part of human-computer interaction, takes user needs into account at every stage of product development. It is an important step and a relatively mature design method for building application programs [25]. Its core is to understand users and build their psychological model into system functions, so as to provide customized high-quality nursing services for each user more effectively [26]. In the process of app development, UCD gives priority to the needs of users, which can improve the stickiness and autonomy of users, and bring positive emotional experience to users, so as to meet the needs, preferences, and goals of users, and improve the quality of apps. The World Health Organization recommends that it be integrated into the whole process of an mHealth intervention to ensure the effectiveness of the intervention [27-30]. In recent years, UCD technology has been applied in the development of mHealth apps and has achieved certain effects in the areas of lifestyle intervention for patients with chronic diseases [25]. Studies have shown that most health management apps are not designed with user-centered methods, leading to poor usability and less use of these apps [31].

At present, some apps for GDM health management have been developed by scholars, but few studies have applied mHealth apps to GDM prevention management [32]. As mentioned above, understanding the usage preferences of mHealth apps among people at risk of gestational diabetes is crucial for the development of GDM prevention apps. However, there is no evidence that people at risk of gestational diabetes have a preference for mHealth apps. Qualitative research methods can dig deep into the inner needs of pregnant women. Therefore, the purpose of this study was to use a user-centered qualitative research method to interview pregnant women at high risk of GDM in early pregnancy, in order to explore the needs and preferences of these pregnant women for the functions and design of mHealth apps. The results can provide a reference for the design and development of mHealth apps for the prevention of gestational diabetes, and can help to adjust intervention measures, improve the acceptance and effectiveness of app use by pregnant women, and improve user engagement.

**Methods**

**Study Design**

In this study, the descriptive qualitative research method was adopted and the semi-structured in-depth interview method was used to collect data. According to the purpose of the study, the interview outline was designed based on previous experience and reference to relevant literature. Before the formal interview, 2 pregnant women were preinterviewed, and the interview outline was modified appropriately according to the results of
the preinterview analysis. The formal interview outline has been presented in Textbox 1.

Textbox 1. Semistructured interview guide used in this study.

<table>
<thead>
<tr>
<th>Guide</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Which mobile health management apps have you used before? What features do you find particularly useful in these apps? Or which features are not very useful and could be improved further?</td>
</tr>
<tr>
<td>2. If an app for gestational diabetes prevention is developed to assist your daily health management, what functions do you expect it to have?</td>
</tr>
<tr>
<td>3. What are your requirements for the interface design of an app?</td>
</tr>
<tr>
<td>4. What other requirements that you would like us to consider for developing an app for gestational diabetes prevention?</td>
</tr>
</tbody>
</table>

Participants and Recruitment

The maximum difference sampling strategy was adopted to select pregnant women with different age, occupation, parity, and gestational diabetes risk factors as far as possible for the purpose of sampling, so as to ensure diversity among the respondents. From July to September 2021, in the obstetrics clinic of a third-class hospital in Beijing, a researcher publicized the project to pregnant women, who made an appointment to establish routine health records, and invited pregnant women to share their experiences and opinions in depth in the form of face-to-face interviews. The inclusion criteria were as follows: (1) at least one risk factor for gestational diabetes (eg, advanced maternal age, overweight or obesity, family history of diabetes, history of GDM, history of macrosomia, and impaired fasting glucose); (2) gestational age <12 weeks; (3) experience of using sports apps, maternal and child health apps, or other health management apps; (4) good communication skills; and (5) informed consent and voluntary participation in this study. The exclusion criteria were as follows: (1) age <18 years and (2) presence of mental disorders. The interviews in this study were stopped when no new topics appeared, that is, the data were saturated. A total of 16 pregnant women were interviewed, and 1 of the women dropped out because she had something else to do halfway. Finally, 15 women participated in semistructured interviews (participant #1 to participant #15). The mean age of the respondents was 32 (SD 3.44) years, and the mean gestational age was 6.8 (SD 0.74) weeks. There were 12 first-time pregnancies and 3 second-time pregnancies. One of the pregnant women was a stay-at-home mother, and the other 14 were from different occupations. Six had 2 or more risk factors, and the remaining 9 had only 1 risk factor. Among the 15 pregnant women, 1 was of Hui nationality and the other 14 were of Han nationality. With regard to the education level, 3 had junior college or below education, 10 had undergraduate education, and 2 had graduate or above education. In terms of cost, 1 of the 15 pregnant women was self-paid, while the rest were covered by medical insurance.

Data Collection

Face-to-face interviews were conducted. Each interview was conducted in a quiet outpatient lounge with no third person to disturb. The interview duration was 20-40 minutes. Before the interview, interviewers introduced to interviewees the definition, harm, and intervention status of gestational diabetes. The purpose and significance of the interviews were also informed. Interviewees were told that their privacy would be protected by the researchers. The interview was recorded with the consent of the interviewees. The interview was conducted according to the interview outline. During the interview, the interviewers carefully listened to the statements; appropriately responded to them with questioning, repetition, clarification, response, and summary; encouraged participants to fully express their ideas; avoided inductive questioning; and timely recorded the key information of the interview. At the same time, they paid attention to observe and record the interviewees’ nonverbal information, such as a pause, a smile, body language, and mood change. After each interview, a reflective diary was written to reflect on the problems in the interview and correct them in the next interview. After the interview, the interviewees were thanked for their participation and were informed of the possibility of contacting them again for further information.

Data Analysis

Data collection and analysis were conducted simultaneously. After each interview, interviewers listened to the original materials repeatedly. The materials were transcribed within 24 hours. Uncertainties were clarified in time in combination with on-site notes. The traditional content analysis method was used for data analysis. The final transcribed text was merged into a single text by topic. Interviewers read the text several times to get a sense of the whole text. Selection criteria were determined according to research objectives and research questions. Based on this standard, the text content was classified, and meaningful statements were extracted and coded. Interviewers read and analyzed the semantic units carefully, and distinguished and summarized the theme. Researchers also looked for the relationship between subjects, and formed a theme group. This cycle continued until saturation (no new themes or subthemes were present) [33].

Quality Control

Before the interviews, the researchers received training in qualitative research, read a large number of relevant literature and books, and learned the analytical methods of qualitative research. The interviewers were involved in the obstetrics clinic as student nurses. As a research tool, researchers always remained neutral. Interviewers truthfully recorded the information provided by interviewees and analyzed their body language and facial expressions. Two researchers with training independently analyzed and discussed the data until the coding information reached a consensus. In the process of data analysis, researchers paid attention to the use of suspension to avoid interference caused by researchers. This study has been reported according to the requirements of the Consolidated Criteria for
Ethical Considerations

Before each interview, the interviewee was given an explanation on the research objectives, methods, expected benefits, and potential risks. Interviewers informed interviewees that relevant information would be strictly confidential. Interviewees could choose to accept or refuse participation in the study, and they could withdraw from the study at any time during the interview. During the interview, interviewees were told that they could refuse to answer any questions that they did not want to answer. Interviewees voluntarily participated in the interviews. In addition, interviewees were told that the content would be used only for scientific research. The interview information was coded, and the researchers did not compromise the privacy of the interviewees.

Ethics Approval

This study was approved by the Ethics Committee of Capital Medical University (batch number: 2019SY037) and the Ethics Committee of the hospital conducting the interviews (batch number: 2019-P2-204-02).

Results

Design

Theme 1: User-Friendly Interface Design

Pregnant women expected clear logic between modules on the app interface and human-computer interaction.

- Apps should have logic. What are the second-level interface and third-level interface after opening app? This hierarchy and interface framework should be clear. Make sure the entry and return routes are clear, and then the modules are clear. [Participant #1]
- Categorize weight management, diet and exercise so that they can be easily seen on the home page. As clear as the app module of hospital appointment. Don't make me look for it, because it's too hard to look for it. A lot of apps these days are really annoying. [Participant #4]
- Hopefully the app doesn't lag, is smooth to use and doesn't have too many ads. [Participant #6]
- App can be divided into modules, like what's this, what's that. And when it updates content, don't always change the location. [Participant #7]
- The app guidelines should be clear. For example, interface modules can be divided into early pregnancy, middle pregnancy and delivery, which are mainly practical and simple. [Participant #10]

Theme 2: Rich Functionality

Pregnant women hoped that the health management apps they use would cover the common functions of blood glucose management during pregnancy, so as to minimize the use of multiple apps.

- I've always fantasized about an app that recommends foods I love. I don't have to think about matching my diet myself. It's all done for me, so just follow the app. But only if I can choose what I like to eat, or if the app removes several ingredients I don't like to eat from the recipe. It would be nice if the app had a search function. For example, if I want to eat strawberries, I can see what the glycemic index of strawberries is and whether it's recommended for me to eat them. [Participant #1]
- I find it convenient to have recommended recipes in the app so I don't have to think about whether the food is edible or not. It's not realistic for me to follow the recipes exactly, but I'm free to mix and match, as long as the total calories are right, and I think it's better. [Participant #4]
- In addition, some pregnant women also expressed a need for exercise customization.
  - I hope the app can let me choose the exercise I want to do every day. For example, today I want to do yoga, and the app can calculate how long I need to do yoga according to my current weight and calorie intake to reach the goal of burning all calories. [Participant #4]

Theme 2: Accurate Information Support

As the pace of work and life is accelerating, some pregnant women stated that popular science articles recommended by mHealth apps should have attractive titles, be relevant to them, and be more accurate.

- At my age, when I have to juggle work and family, I really don't have much free time. For popular science articles, the content is already boring, and if the volume is any longer, users will have no interest or time to read it. [Participant #1]
- I hope that the app can set up a function of a prenatal assistant to tell me what I should do at different gestational weeks and whether I should have an empty
stomach, etc. This is my favorite feature. [Participant 

The app should recommend relevant contents for pregnant women based on their individual weight and test values. I think this will be better; because I think people are extremely busy nowadays, especially when they have children. [Participant #4]

We hope that the app can push suitable exercise types for pregnant women according to their gestational weeks. [Participant #7]

If it’s a video, it can be put in the appropriate module and I’ll pay attention to it when I get to that stage of pregnancy. For example, if it’s about breastfeeding, it’ll be put in the third trimester and I’ll click on it when I’m in labor. [Participant #10]

Only when the title of an article is clear and relevant to me will I read it. For example, if I’m eight weeks pregnant, what do I need to be aware of at this stage? I might click on it. But if it’s about what to eat and control, I’m probably not going to read it. [Participant #11]

**Theme 3: Interactive Design**

Pregnant women hoped that mHealth apps have the function of setting daily calorie and exercise goals. After the pregnant women input their own diet, exercise, and blood sugar monitoring data, the system should provide immediate feedback and professional feedback according to their health management standards, so that the pregnant women can make self-adjustments to achieve their daily management goals.

I hope the app can record the number of steps on the day, and tell me whether the amount of exercise today meets the standard, so as to give me an evaluation. [Participant #1]

If I have a problem with my blood sugar, I may check my blood sugar every once in a while and record the result of each time on the app. Then the application platform will give feedback based on my data and tell me what I need to pay attention to in the next step. I think it would be nice to have that kind of feedback. [Participant #4]

The app can score me according to my exercise level every week, and then give me an adjustment plan, which forms a closed loop from input, scoring and feedback. [Participant #10]

**Theme 4: Practical Tool Support**

Pregnant women hoped that mHealth apps can be used as tools to assist them in self-management and that these apps can record their steps to help them know whether their exercise meets the standard, or can remind and motivate them to exercise.

Although I am in a first-tier city like Beijing, there are few pregnant women around me who do yoga or swim. Especially when you’re expecting your first child, families are very cautious. The elders in the family are so ingrained in their beliefs that they may not be comfortable with their children doing yoga during pregnancy, and most would probably prefer to stick to walking during pregnancy. Therefore, it is sufficient for the app to record the number of steps of motion. [Participant #5]

For example, I can punch in the app to record my steps today. [Participant #7]

In addition, pregnant women hoped that mHealth apps have the function of weight recording to show the trend of their own weight change and that these apps can automatically compare the actual weight increase of pregnant women with the recommended weight increase range to determine whether the weight gain is reasonable, so as to help pregnant women manage their weight during pregnancy.

The app needs to have the function of weight graph, through which we can see the recent trend of pregnant women’s weight. I remember when I was pregnant with my first child, there was an app that told me how much weight I needed to gain in my current gestational week. [Participant #4]

I am using an app for weight monitoring. I need to input my weight into the app every day to observe the change of my weight. [Participant #9]

In addition, some pregnant women hoped that mHealth apps can be used as practical tools to assist doctors in pregnancy management.

The app itself is also a tool to help doctors manage pregnancy health for pregnant women. Since everyone is different, doctors will definitely need to manage pregnancy according to their different physical conditions. It would be better if the app could be linked to hospital records and checklists. [Participant #6]

**Theme 5: Visual Presentation**

Some pregnant women stated that they do not have time to watch live broadcasts because of the fast pace of life. They hoped that the information on mHealth apps can be presented in the form of cartoon pictures, risk assessment forms, recorded videos, and short videos, so as to obtain relevant information more quickly.

The recommended article can be in the form of text with cartoon pictures, or it can be one of those risk self-assessment scales for pregnant women to rate their recent status. If the risk score is higher than a certain point, the woman has gestational diabetes. [Participant #10]

The live broadcast lasts a long time, and I may not be able to listen to it, because the pace of life is fast now, and life is so busy every day. Let alone live broadcast, I may skip to watch the recorded broadcast. [Participant #5]

I think recording is better than live broadcasting. Medical staff can put the video of the lecture in the corresponding module. If I had time, I would listen. [Participant #10]
I may not have time to watch the health education live broadcast, if there is a replay, I will watch it. It's better to put subtitles on the video because it's slow and I tend to skip to the subtitles or just read the document. [Participant #6]

If experts' lectures are made into some short videos, I may watch them at any time. [Participant #4]

Theme 6: Convenient Professional Support
Some pregnant women hoped that mHealth apps can provide the online consultation function of experts, so that it is more convenient to contact experts and solve some minor problems that are not too urgent, in order to avoid the cost of time and energy from the round trip to the hospital.

It's more difficult for pregnant women to get to the hospital. If I have any minor questions, I would like to consult one of the experts while they are giving an online lecture. I hope the app can provide free online consultation. [Participant #1]

Pregnant women may have some simple problems, but it is not convenient to come to the hospital for registration. It would be better if the app could provide online consultations with doctors. Doctors can choose to reply to pregnant women's information when they have time, there is no rush to reply immediately. [Participant #4]

It would be better if pregnant women had some problems that did not need to go to the hospital and could be answered online at a lower cost than going to the hospital. [Participant #9]

Theme 7: Peer Support
Some pregnant women hoped to set up a function like WeChat Moments or a forum, in order to facilitate access to other pregnant women who have encountered problems and facilitate the exchange of experiences between pregnant women, so as to better deal with the problems of pregnancy.

When I was first pregnant, I went to the forum to see what other people were like when they were pregnant. Many people would share their early pregnancy experiences on the forum, such as problems with their checklists. [Participant #1]

If I have some problems, I will look at other users' posts to see how they solve the problems. For example, I have a tummy ache at the early stage of pregnancy, so I will search for users who have the same experience on the Internet. If most users say this is normal, it means that tummy ache is ok. [Participant #8]

Theme 8: Reasonable Reminder Function
Pregnant women hoped to set reminders according to their own needs. Excessive reminders can disturb pregnant women.

We may not always want to record our diet every day. I hope the app can remind me when I forget to fill in. [Participant #7]

I hope the app has a reminder function, such as when I should do activities, when I should drink water, timely reminding me will be better. [Participant #10]

I think it is still necessary to set reminders, because the pace of work and life is fast now, pregnant women cannot remember everything every day. [Participant #14]

However, some pregnant women believed that an app’s reminder function would cause some trouble, so they had little demand for the reminder function.

Because there are too many reminders in the app, I don't set any reminders in all my apps. I turn them off. If I want to keep track of something important, I'll set up a reminder myself. [Participant #9]

Theme 9: Appropriate Maternal and Infant Auxiliary Function
Some pregnant women wanted to add mother-baby support tools to mHealth apps to make it through the pregnancy better.

I hope the app can tell me the growth of my baby, so that I can know the development of the baby. In addition, I hope the app can tell me what changes a pregnant woman will have during pregnancy and what changes are normal. [Participant #1]

I think the app can add a function to record the gestational weeks. [Participant #2]

I have heard that some apps can record contractions and fetal movements. Is it possible to add these functions? [Participant #3]

I think there should be a keyword search function, I can query relevant articles, just like Baidu, I want to know can be found in it. [Participant #9]

Theme 10: Differentiated Incentive Function
Different pregnant women had different views on the redemption function of points. Some thought that it would be better if the points can be exchanged for items they need, while some thought that it has no incentive effect on them.

It is possible to get points by checking in, posting, reading articles, ranking points, and exchanging points for small things, etc. Some people may prefer this function. [Participant #5]

I think it would be nice if I could exchange my accumulated points for something to use during pregnancy or in the future. [Participant #6]

I don't like those things very much. They are not very useful. I sometimes check to see how many points I have, but I don't think the points are of any use to me. [Participant #7]

Discussion
Principal Findings
From the perspective of users, this study discussed the views and function requirements of pregnant women at high risk of GDM with regard to the development of mHealth apps for
preventing GDM, which can provide a reference for app developers to develop apps in line with users’ usage habits in the future. In this study, pregnant women hoped that when using apps for management, they could receive personalized health management strategies according to their different stages of pregnancy and could receive immediate feedback on their daily management, helping them make adjustments. When pregnant women encountered difficulties in self-management, they hoped to obtain expert and peer resources in the apps to provide them with social support. In addition, pregnant women hoped that developers and professionals would consider the need for quick and accurate information acquisition when designing mHealth apps and would develop user-friendly apps. They stated that the content should be visual and attractive, and should provide both valuable and accurate information to meet the needs of information support.

Comparison With Prior Work

This study found that it is of great significance to formulate corresponding personalized management strategies for pregnant women based on the results of their health assessment, which is similar to the results of previous studies [35-37]. A research report pointed out that women were dissatisfied with the limited response of apps, and hoped that the apps could specifically analyze the characteristics of users and provide different solutions [38]. The biggest difference between personalized apps and other apps is that personalized apps can provide personalized services for pregnant women in different stages of pregnancy [39]. Studies have shown that personalized features in apps could improve users’ compliance with lifestyle interventions, which is one of the most common intervention strategies in mHealth apps [40].

Interactive function is considered to be the most popular function in mobile medical apps [41]. The biggest advantage of interactive apps is that they can formulate targeted behavior change plans for users, continuously monitor the behavior characteristics of users, and then give targeted feedback according to the implementation of users’ behaviors, which can motivate users to complete behavior changes [42]. Interactive design strengthens the connection between the user and the mobile app, providing psychological support for pregnant women [43]. In this study, pregnant women hope that the apps could compare the health information of pregnant women with the recommended behavioral goals, and provide feedback instantly for pregnant women to help them make timely behavioral adjustments, which is consistent with previous research results [44-46]. Some studies have used telephone or WeChat groups to intervene in pregnant women at high risk of GDM [47,48]. However, it is difficult to achieve instant feedback between doctors and patients, and the information of telephone and WeChat groups is also easily forgotten and covered. Therefore, it is important to design an app to prevent gestational diabetes. There is also a study using an app that counts low glycemic index function to provide dietary guidance to obese pregnant women, but this app has a single function and is designed for dietary guidance only, which cannot achieve the goal of interaction, resulting in an insignificant intervention effect [49].

Due to time constraints, most working pregnant women hoped apps would accurately push information related to themselves. Previous studies have found that providing users with personally relevant information was an important factor in promoting the use of mHealth programs [50]. In addition, Naughton et al [51] found that pushing personalized and relevant information could increase the value of contents and prevent users from falling out of mHealth management. In a qualitative study of Saudi women, it was found that due to social and cultural restrictions, some obese women could not do enough physical exercise outdoors [52]. Interviewees hoped that apps could recommend exercise methods that meet actual conditions for users and provide advice according to users’ preferences.

Most pregnant women hoped that information could be in visual forms, such as illustrations and videos, to increase attractiveness. Presenting content in a visual way can help women obtain more information in a short period of time to meet women’s needs for pregnancy knowledge. Previous studies have pointed out that multimedia could adapt to the learning styles of different individuals, enable pregnant women to take the initiative in learning, and encourage pregnant women to actively study [53,54]. Some studies have pointed out that the memory acquired by watching videos was stronger than that acquired by browsing text or pictures [55]. This study found that in terms of the video presentation form, pregnant women were more inclined to short videos or recorded videos. Therefore, when designing apps, developers should use a visual method as much as possible to present a variety of health information and educational resources for pregnant women to meet their needs for information support.

During the daily management of pregnancy, pregnant women will obtain decision-making information from different channels. Studies have shown [56] that pregnant women tend to trust professional decision support. Professional support provided by medical staff can meet the needs of pregnant women in pregnancy health care, which relieves anxiety and uncertainty of women during pregnancy [57,58] and helps pregnant women establish a healthy lifestyle [59]. When professionals cannot provide timely and effective feedback, pregnant women usually seek help from relatives and friends or through the internet. However, due to the poor quality of advice from various parties, pregnant women often get confused and make wrong decisions, which is not conducive to health management [60]. This study found that pregnant women hoped to obtain professional feedback from apps when they encountered problems or in the process of daily health management. This is similar to previous studies by Edwards et al [61] and Lau et al [43]. Previous studies have shown that when patients were unable to communicate effectively with experts, they felt bored and useless to use an app, resulting in a low utilization rate [36]. Therefore, when developing apps in the future, professional support functions should be added. In addition, when apps are actually used to manage the health of pregnant women, managers should coordinate the workload distribution of medical staff and include answering questions from pregnant women into their work scope, providing users with timely and effective professional support.

In addition to professional support, peer support also plays an important role in the formation and maintenance of patients’
self-management abilities. Peers can communicate with each other on their own experience, attitudes, and concepts, giving relevant suggestions to each other. Pregnant women in this study hoped that they could share confusion, emotion, and experience with peers who had similar needs and common goals through an app; thus, these women can obtain information and emotional support. This is similar to the findings of McDonald et al [62]. Studies have shown that peer support could overcome loneliness, powerlessness, and stress in pregnant women and increase maternal self-efficacy. Emotional support and exchange of experiences with peers are more acceptable to pregnant women. In addition, pregnant women can gain motivation from peer support to use an app [63]. Previous studies have found that providing a communication platform for users in a GDM prevention app could allow users to exchange their experiences and deal with problems during pregnancy, motivate users to achieve their expected goals, and increase the frequency of usage [64]. Therefore, apps should provide a module for pregnant women to communicate with their peers, allowing them to learn problem-solving skills through mutual assistance and to enhance psychological support.

Although some pregnant women in this interview study believed that the point reward function was of little significance, research has proven that the introduction of game mechanics could increase the user’s sense of achievement and could cause the user to put more effort into accomplishing goals. Cafazzo et al [65] incorporated a gamification mechanism when designing a diabetes blood sugar management program. When young patients regularly checked their blood glucose, they could obtain corresponding points. These points could be redeemed for iTunes codes to purchase music and applications. The frequency of blood glucose monitoring increased by nearly 50% in patients using the apps. Ekezie et al [64] provided a virtual map for gestational diabetes mellitus patients in his app to improve the interest of patients’ exercise, and set a step ranking list to encourage users to achieve exercise goals through competition, so as to increase user efficiency. It is suggested that when developing gestational diabetes prevention apps in the future, the gamification and reward mechanisms should be improved according to users’ needs. Furthermore, more attractive methods should be set up to better assist patients in performing self-management, as well as raise their enthusiasm for using these apps.

Strengths and Limitations

Previous research has mostly focused on the user’s experience and evaluation of developed apps after using them, and few users have participated in the design of apps, leading to poor user stickiness. In this study, we conducted deep interviews on the preferences and functional needs of mHealth apps among people at risk for gestational diabetes, and the findings are important to guide the development of future interventions and other pregnancy health apps. In addition, the strength of this study is that the data analysis was performed by two researchers and information was continuously validated during the data analysis process. One of the researchers was the interviewer and the other was not involved in the interview. Therefore, in the process of data analysis, the two researchers had a different understanding of the existing data. Then they discussed and finally reached a more objective and unified opinion, and provided the analysis results to the participants to make their opinions expressed fully and correctly.

This study has several limitations. First, this study only interviewed 15 pregnant women in the same hospital in Beijing, China. Regardless of the fact that the sample size of this study reached saturation, the method of maximum difference sampling was adopted to select the research objects as much as possible. However, no interviewees from rural groups were included; thus, the interview results may not reflect the diverse needs of all pregnant women, making it difficult to generalize the findings to women in rural areas or different cultural situations. Before the formal interview, interviewers trained and practiced to improve their interviewing skills. However, as interviewers were conducting interviews for the first time, their interviewing experience and skills were limited, which may have affected the comprehensiveness of data collection. In addition, interviewers conducted interviews as medical interns, which made some interviewees reluctant to express their true feelings and needs, leading to a potential impact on data collection.

Conclusion

mHealth apps can provide new tools for the health management of people at risk for gestational diabetes. Before the development of a gestational diabetes prevention app, this study deeply collected the views and functional requirements of mHealth app development from gestational diabetes risk groups. The results can provide valuable information for the future development of mHealth apps related to the prevention of gestational diabetes that meet users’ needs, and can provide a reference for the design of intervention content in the future. In addition, the findings can provide a reference for the development of other types of apps during pregnancy.

Acknowledgments

The authors sincerely thank all the interviewees for their active participation in this study. This study was supported by the National Natural Science Foundation of China (project number 71904133).

Authors’ Contributions

BD performed data collection, performed data analysis and interpretation, and drafted the manuscript. ZL assisted in qualitative data analysis and interpretation, and co-drafted the manuscript. BG participated in manuscript review. WL conceived and designed the study and was responsible for revising the manuscript. All authors read and approved the final manuscript.

https://formative.jmir.org/2022/7/e36392
Conflicts of Interest
None declared.

References


Abbreviations

GDM: gestational diabetes mellitus
mHealth: mobile health
UCD: user-centered design

©Beibei Duan, Zhe Liu, Weiwei Liu, Baohua Gou. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.07.2022. 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.
Health-Related Quality of Life Among Members Using an On Demand Behavioral Health Platform: Pilot Observational Study

Emily Shih\textsuperscript{1}, PhD; Brandon S Aylward\textsuperscript{1}, PhD; Sarah Kunkle\textsuperscript{1}, MPH; Grant Graziani\textsuperscript{1}, PhD

Ginger, San Francisco, CA, United States

Corresponding Author:
Emily Shih, PhD
Ginger
116 New Montgomery St. Suite 500
San Francisco, CA, 94105
United States
Phone: 1 6282079974
Email: eshih@ginger.io

Abstract

Background: Despite the well-known adverse health conditions and negative economic outcomes associated with mental health problems, accessing treatment is difficult due to reasons such as availability and cost. As a solution, digital mental health services have flooded the industry, and new studies are quickly emerging that support their potential as an accessible and cost-effective way to improve mental health outcomes. However, many mental health platforms typically use clinical tools such as the Patient Health Questionnaire-9 (PHQ-9) or General Anxiety Disorder-7 (GAD-7). Yet, many individuals that seek out care do not have clinical symptomatology and thus, traditional clinical measures may not adequately capture symptom improvement in general well-being. As an alternative, this study used the health-related quality of life (HRQoL) tool from the Centers for Disease Control and Prevention “Healthy Days” measure. This subjective measure of well-being is an effective way to capture HRQoL and might be better suited as an outcome measure for treatments that include both clinical and subclinical individuals.

Objective: The purpose of this study was to describe changes in HRQoL in clinical and subclinical members assessing virtual care and to examine the association between text-based behavioral coaching and virtual clinical sessions with changes in HRQoL.

Methods: A total of 288 members completed the 4-item HRQoL measure at baseline and at 1 month following use of the Ginger on demand behavioral health platform. Baseline anxiety and depression levels were collected using the GAD-7 and PHQ-9, respectively.

Results: Members completed on average 1.92 (SD 2.16) coaching sessions and 0.91 (SD 1.37) clinical sessions during the assessment month. Paired samples t tests revealed significant reductions in the average number of unhealthy mental health days between baseline (mean 16, SD 8.77 days) and follow-up (mean 13.2, SD 9.02 days; \(t_{287}=5.73; P<.001\)), and in the average number of days adversely impacted (mean\textsubscript{baseline} 10.9, mean\textsubscript{follow-up} 8.19; \(t_{287}=6.26; P<.001\)). Both subclinical members (\(t_{103}=3.04; P=.003\)) and clinical members (\(t_{183}=5.5; P<.001\)) demonstrated significant improvements through reductions in adversely impacted days over a month. Clinical members also demonstrated significant improvements through reductions in unhealthy mental health days (\(t_{183}=5.82; P<.001\)). Finally, member engagement with virtual clinical sessions significantly predicted changes in unhealthy mental days (\(B=-0.96; P=.04\)).

Conclusions: To our knowledge, this study is one of the first to use the HRQoL measure as an outcome in an evaluation of a digital behavioral health platform. Using real-world longitudinal data, our preliminary yet promising results show that short-term engagement with virtual care can be an effective means to improve HRQoL for members with subclinical and clinical symptoms. Further follow-up of reported HRQoL over several months is needed.

(JMIR Form Res 2022;6(7):e35352) doi:10.2196/35352

KEYWORDS
behavioral coaching; mental health; telehealth; Healthy Days; clinical care; behavior; coach; quality of life; platform; tool; pilot study; observational; health-related quality of life; virtual care; association; text-based; outcome; evaluation
Introduction

Nearly 1 in 5 adults in the United States (51.5 million people) experience mental health issues [1]. The World Health Organization estimates that anxiety and depression alone cost the global economy US $1 trillion dollars each year in lost productivity, absenteeism, and medical costs [2]. Mental health issues have been exacerbated with the recent COVID-19 pandemic and underscore a critical moment of global need [3,4]. A recent meta-analysis found the global prevalence of diagnosable anxiety and depression during the pandemic was 27% and 28%, respectively [5]. Even among those with subclinical symptoms, nearly half of adults in the United States have reported symptoms of anxiety or depression during this time [6]. Timely intervention for those with subclinical symptoms is just as important to prevent development of more serious symptoms requiring more costly treatment.

Despite the well-known adverse health conditions and negative economic outcomes, accessing treatment for common mental health problems is difficult [7]. The demand for mental health services has outpaced the availability of qualified mental health professionals. A recent survey found that 1 in 4 individuals with depression or anxiety lack access to care or have unmet mental health needs [8]. In addition, long wait lists, high out-of-pocket expenses, and transportation burdens all continue to serve as barriers to receipt of effective services [9,10]. There is a growing need for scalable mental health solutions that increase both the availability of professionals and access to care for common mental health conditions. This is particularly important with the recent increase of mental health issues during the pandemic. Digital mental health services have flooded the industry, and new studies are emerging that support their potential to serve as cost-effective ways to manage anxiety and depression [11,12]. This type of support can even be beneficial for individuals who may be at risk for but do not yet experience clinically significant symptoms [13].

Many mental health platforms typically use clinical tools such as the Patient Health Questionnaire-9 (PHQ-9) or General Anxiety Disorder-7 (GAD-7) for assessing initial and treatment outcomes of depressive and anxiety symptoms, respectively. As behavioral coaching focuses on goal-oriented behavior and typically targets those with subclinical symptomatology, traditional clinical measures may not adequately capture symptom improvement in general mental health and well-being. State and federal health agencies have supported the population surveillance of health-related quality of life (HRQoL), which is a multidimensional concept that examines overall health related to perceived physical and mental health as well as daily functioning [14,15]. One common HRQoL tool is the Centers for Disease Control and Prevention (CDC) “Healthy Days” measure that asks about self-rated general health, physical health, mental health, and activity limitations over the past 30 days. This subjective measure of well-being is an effective way to capture HRQoL and might be better suited as an outcome measure for treatments that include both clinical individuals and individuals with symptoms not meeting clinical thresholds [16]. Yet, few studies have used this measure when evaluating digital behavioral health platforms. Financially, Humana found that the cost of each reported unhealthy day is equivalent to 10 hospital admissions per thousand patients, with a potential increase of US $15.64 per member per month in medical costs for each unhealthy day [17]. This highlights the potential long-term savings that could result from interventions targeting individual HRQoL. A previous health coaching study has already demonstrated significant reductions in reported unhealthy days among participants [18].

The purpose of this study was to examine self-reported HRQoL among members using an on demand digital health platform and the association of short-term text-based behavioral health coaching and virtual clinical sessions with healthy days over time. To that end, the study will describe baseline characteristics of members in terms of reported unhealthy days and changes over 1 month, describe changes in unhealthy days as a function of baseline anxiety and depressive symptoms, and examine the association between member engagement and changes in unhealthy days.

Methods

Participants

Participants were members who had access to the Ginger on demand behavioral health platform as part of their employer or health plan benefits. Internal clinical protocols include exclusionary criteria where self-directed telehealth is likely not appropriate and where more specialized and urgent psychiatric services are required (eg, active suicide ideation or active high-risk self-harm behavior; see Kunkle et al [19] for exhaustive list). This study included Ginger members 18 years or older who completed the baseline Healthy Days measure between November 2020 to November 2021 and who first accessed care within 1 month of completing their Healthy Days baseline.

Procedures

The Ginger platform provides members with access to virtual behavioral health coaching, teletherapy, telepsychiatry, and self-guided content and assessments, primarily via a mobile app platform. After downloading the mobile app, members can start texting with a behavioral health coach within minutes of requesting to connect. Ginger coaches are full-time employees who have an advanced degree in a field related to mental health or have accredited coach certification. While many members are solely engaged with text-based coaching services, some will request or require escalation to clinical services (teletherapy or telepsychiatry) depending on preference or clinical severity. When members are escalated to therapy or psychiatry, they may continue working with a coach provided they also seek additional specialized care concurrently. Additional detail regarding Ginger can be found in prior publications [19,20].

The Healthy Days measure was administered to members 4 times across the span of 4 months (once per month). Data were collected externally using the Survey Monkey platform. Only responses from survey items pertaining to the number of unhealthy mental health days and impacted days were of focus for this study. The PHQ-9 and GAD-7 were typically completed

https://formative.jmir.org/2022/7/e35352
at intake within 1 month of the Healthy Days baseline assessment.

Measures

The CDC Healthy Days measure contains four items: (1) “Would you say that in general your health is excellent, very good, good, fair, or poor?” (2) “Now thinking about your physical health, which includes physical illness and injury, for how many days during the past 30 days was your physical health not good?” (3) “Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?” and (4) “During the past 30 days, for about how many days did poor physical or mental health keep you from doing your usual activities, such as self-care, work, or recreation?” (referred to here as impacted days). For this study a change variable was calculated by subtracting reported unhealthy scores from time 1 from scores from time 2, where positive values indicate an increase in unhealthy days, whereas negative values indicate a reduction in unhealthy days.

The PHQ-9 is a 9-item self-report questionnaire that assesses the frequency and severity of depression symptomatology within the previous 2 weeks. Each of the 9 items is based on the Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition; DSM-IV) criteria for major depressive disorder and are scored on a 0 (not at all) to 3 (nearly every day) scale. Items include “Little interest or pleasure in doing things” and “Feeling down, depressed, or hopeless.” Total scores can range from 0 to 27 with higher scores indicating more depressive symptoms. A score of 10 was used as the clinical threshold [21].

The GAD-7 is a valid brief self-report tool to assess the frequency and severity of anxious thoughts and behaviors over the past 2 weeks. Each of the 7 items are based on the DSM-IV diagnostic criteria for generalized anxiety disorder and are scored on a 0 (not at all) to 3 (nearly every day) scale, with total scores ranging from 0 to 21. Items include “Feeling nervous, anxious, or on edge” and “Not being able to stop or control worrying.” Consistent with existing literature, a score of 10 was used as the clinical threshold [22].

Member engagement with Ginger services was quantified as the number of coaching and clinical sessions. Coaching sessions were operationalized as the number of unique days where both members and coaches each sent at least 5 text messages. Ginger coaching is an on demand text-based service, and the operationalization of a “text-based coach session” has not been predetermined in the literature. As such, our threshold was decided based upon internal work that highlighted approximately 5 texts each way as the number of text messages needed to capture a productive conversation between members and their coaches. Clinical sessions were operationalized as the number of completed video sessions with a clinician.

Statistical Analysis

Analyses were conducted using RStudio (version 1.4.1717; RStudio, PBC). Data were first screened for outliers and normality. Descriptive statistics were used to describe baseline member characteristics. For changes in reported unhealthy days, paired sample t tests were used. Next, members were divided into groups as a function of clinical thresholds using the PHQ-9 and GAD-7 scores at intake (ie, clinical vs subclinical). Additional paired sample t tests were performed to evaluate member differences in responses between time 1 and time 2 for clinical and subclinical groups separately. A Benjamini-Hochberg correction was used to adjust for multiple comparisons [23]. Finally, scatterplots suggested a linear trend between member engagement and changes in unhealthy days. As such, multiple linear regressions were performed to examine the association of member engagement (ie, coaching and clinical sessions) with changes in the number of unhealthy days. Baseline Healthy Days scores and the number of prior engagement levels were entered as covariates. All continuous variables were standardized for interpretability.

Ethics Approval

This is a secondary analysis of pre-existing deidentified data. The authors do not have access to participant identifying information and do not intend to recontact participants. Ginger’s research protocols and supporting policies have been reviewed and approved by Advarra’s institutional review board (Pro00046797) in accordance with the US Department of Health and Human Services regulations at 45 CFR 46.

Results

Descriptive Statistics

A total of 1496 members completed the Healthy Days measure at time 1 (intake), 351 (23.5%) members at time 2 (~30 days following intake; mean 31.9, SD 1.48 days), 114 members at time 3 (~60 days following intake), and 37 members at time 4 (~90 days following intake). The current analyses examined only members who had completed surveys at both time 1 (intake) and at time 2 (N=288). Data were missing at random for all primary outcome variables (r>.70 and r<1.54; P>.12). Potential reasons for earlier drop-offs that should be taken into consideration when interpreting our results include members having achieved their coaching goals, members no longer interested in care, and members engaged at a monthly cadence and returned after the study evaluation month was finished. Demographic information about members was provided by employers but contained missing data. Of members in the analytical sample, 82 (28.5%) members were between the ages of 18-34 years, 96 (33.3%) members were 35 years of age or older, and 110 (38.2%) members did not have age reported. Regarding gender identity, 125 (43.4%) members identified as female, 33 (11.5%) as male, 14 (4.9%) as other, and 116 (40.3%) did not have gender reported.

Descriptive statistics for the primary variables are presented in Table 1. Members, on average, completed 1.92 (SD 2.16, range 0-12) coaching sessions and 0.91 (SD 1.37, range 0-5) video sessions with a clinician within a single month. A total of 179 (62.2%) members engaged exclusively with text-based coaching (no clinical sessions). Subclinical depression and anxiety levels were reported in 104 (36.1%) members, whereas 184 (63.9%) members reported clinical levels of depression or anxiety. Of members in the analytical sample, 71% (n=205) at time 1 and 77% (n=223) of members at time 2 reported feeling “good” or better in response to the question “Would you say that in general
your health is excellent, very good, good, fair, or poor?” (includes members who reported feeling “very good” and “excellent”). Bivariate correlations among the primary variables are presented in Figure 1. Of note, the number of unhealthy mental health days was positively correlated with the number of impacted health days at each respective time point ($r=0.62$ at time 1, $r=0.65$ at time 2; $P<.001$).

Table 1. Descriptive statistics among primary variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Values, mean (SD)</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical health (time 1)</td>
<td>5.1 (7.8)</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>Mental health (time 1)</td>
<td>16.0 (8.8)</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>Impacted health (time 1)</td>
<td>10.9 (9.6)</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>Physical health (time 2)</td>
<td>5.6 (8.3)</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>Mental health (time 2)</td>
<td>13.2 (9.0)</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>Impacted health (time 2)</td>
<td>8.2 (8.5)</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>Coaching sessions</td>
<td>1.9 (2.2)</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Clinical sessions</td>
<td>0.9 (1.4)</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Depression score (PHQ-9&lt;sup&gt;a&lt;/sup&gt;)</td>
<td>11.3 (6.1)</td>
<td>1</td>
<td>27</td>
</tr>
<tr>
<td>Anxiety Score (GAD-7&lt;sup&gt;b&lt;/sup&gt;)</td>
<td>9.8 (5.7)</td>
<td>0</td>
<td>21</td>
</tr>
</tbody>
</table>

<sup>a</sup>PHQ-9: Patient Health Questionnaire-9.

<sup>b</sup>GAD-7: General Anxiety Disorder-7.

Figure 1. Correlations among primary variables. Note: Insignificant correlations where $P>.05$ are marked. Corr: correlation.

Pre-Post Changes in Reported Unhealthy Days

Members reported on average nearly 3 fewer unhealthy mental health days (mean $-2.71$, SD $8.03$) between baseline and 1 month later. Of the analytical sample, 61% (n=175) of members reported an improvement in unhealthy mental health days, whereas 39% (n=113) reported no improvement or an increase in unhealthy mental health days. Paired sample $t$ tests were...
performed to evaluate differences in member Healthy Days responses at time 1 compared to time 2 (Figure 2) across all continuous items. Results showed no significant improvements in unhealthy physical health days between time 1 (mean 5.08, SD 7.78 days) and time 2 (mean 5.60, SD 8.25 days; $t_{287}=-1.25; P=.21$). However, results showed significant improvements in unhealthy mental health days between time 1 (mean 16, SD 8.77 days) and time 2 (mean 13.2, SD 9.02 days; $t_{287}=5.73; P<.001$), as well as significant improvements in adversely impacted days between time 1 (mean 10.9, SD 9.60 days) and time 2 (mean 8.19, SD 8.51 days; $t_{287}=6.26; P<.001$). Given Ginger is a mental health platform and significant changes were only observed for unhealthy mental health days and adversely impacted days, these two outcomes were explored in subsequent analyses.

**Figure 2.** Display of means across the items from the Healthy Days measure at time 1 and time 2 (N=288).

**Comparison of Change in Healthy Days Between Clinical and Subclinical Members**

Subclinical members showed trending reductions in reported unhealthy mental health days between time 1 (mean 9.92, SD 6.78 days) and time 2 (mean 8.44, SD 7.83 days; $t_{103}=1.87; P=.06$, adjusted $P=.06$). Clinical members also showed reductions in reported unhealthy mental health days between time 1 (mean 19.4, SD 7.91 days) and time 2 (mean 16.0, SD 8.52 days; $t_{183}=5.82; P<.001$; adjusted $P<.001$).

Similarly, subclinical members showed significant reductions in reported impacted days at time 1 (mean 5.15, SD 6.64 days) compared to time 2 (mean 3.47, SD 5.3 days; $t_{103}=3.04; P=.003$, adjusted $P=.003$). Clinical members also showed significant reductions in reported impacted days at time 1 (mean 14.2, SD 9.48 days) compared to time 2 (mean 10.9, SD 8.83 days; $t_{183}=5.50; P<.001$, adjusted $P=.001$).

**Member Engagement on Changes in Reported Unhealthy Mental Health Days**

The linear regression model predicting changes in reported unhealthy mental health days was significant ($F_{5,282}=14.6; P<.001$) and accounted for 21% of the variance. No significant main effects of coaching sessions ($B=0.61; P=.19$) were observed. However, there was a significant main effect of clinical sessions ($B=-0.96; P=.04$), where more clinical sessions was associated with a decrease in unhealthy mental health days. The model predicting changes in adversely impacted days was also significant ($F_{5,282}=22.2; P<.001$) and accounted for 28% of the variance. No significant main effects of coaching sessions ($B=0.43; P=.30$) or clinical sessions ($B=-0.40; P=.33$) were observed. Coefficients for both models are presented in **Table 2**.
Table 2. Summary of regression coefficients (N=288).

<table>
<thead>
<tr>
<th>Model 1: Changes in the number of unhealthy mental health days</th>
<th>Beta (SE)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>−2.71 (0.43)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Unhealthy mental health days (baseline)</td>
<td>−3.32 (0.43)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Prior coaching sessions</td>
<td>0.39 (0.47)</td>
<td>.42</td>
</tr>
<tr>
<td>Prior clinical sessions</td>
<td>0.59 (0.47)</td>
<td>.21</td>
</tr>
<tr>
<td>Clinical sessions</td>
<td>−0.96 (0.47)</td>
<td>.04</td>
</tr>
<tr>
<td>Coaching sessions</td>
<td>0.61 (0.47)</td>
<td>.19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 2: Changes in the number of adversely impacted days</th>
<th>Beta (SE)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>−2.75 (0.38)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Unhealthy impacted days (baseline)</td>
<td>−3.87 (0.38)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Prior coaching sessions</td>
<td>−0.30 (0.42)</td>
<td>.48</td>
</tr>
<tr>
<td>Prior clinical sessions</td>
<td>0.10 (0.41)</td>
<td>.80</td>
</tr>
<tr>
<td>Clinical sessions</td>
<td>−0.40 (0.41)</td>
<td>.33</td>
</tr>
<tr>
<td>Coaching sessions</td>
<td>0.43 (0.41)</td>
<td>.30</td>
</tr>
</tbody>
</table>

**Discussion**

**Principal Findings**

This study evaluated the real-world association between digital care utilization in members with both subclinical and clinical symptoms of anxiety or depression. HRQoL at baseline suggested that members were, on average, demonstrating “frequent distress” and reporting more unhealthy mental health days than healthy mental health days (mean 16, SD 8.77 days; 53% of the month). The CDC defines having ≥14 unhealthy mental health days as “frequent distress [24].” Of note, our results also observed a relatively high number of unhealthy mental health days (mean 9.92, SD 6.78 days; 33% of the month) for subclinical members at baseline, highlighting the need for care for those that might not traditionally be recommended for clinical services (eg, individuals who might not have exceeded clinical thresholds using traditional PHQ-9 and GAD-7 assessment surveys). Bivariate correlations revealed a positive association between unhealthy mental health days and adversely impacted days, underscoring the relationship between mental health and daily functioning [25,26]. Overall, members evidenced significant improvements in reported unhealthy mental health days and adversely impacted days over the month. Furthermore, improvements in reported adversely impacted days were significant for both subclinical members and clinical members, and improvements in reported unhealthy mental health days were significant for clinical members. Our results also found that clinical sessions, but not coaching sessions, predicted changes in reported unhealthy mental health days over the month. Taken together, this study offers preliminary descriptive on a valuable but less commonly used outcome measure, specifically in a traditionally understudied but increasing population of individuals seeking out virtual care. The study further supports how virtual care is a promising strategy to meet the growing demand of mental health services.

Not all individuals seeking out care exceeded industry clinical thresholds. Thus, additional outcome measures, such as the Healthy Days measure, are needed to evaluate the effects of digital mental health care beyond clinically focused measures (eg, PHQ-9 and GAD-7). To our knowledge, we are one of the first to use the Healthy Days measure within this population (ie, individuals seeking out virtual mental health care). Overall, members reported a reduction of 2.71 unhealthy mental health days. Extrapolating from the Humana data [16], this would be equivalent to a decrease of 27.1 hospital admissions per thousand patients and a potential cost savings of US $42.38 per member per month. Thus, virtual mental health care can be seen as a low-intensity approach to achieve better health outcomes at lower cost [12,13].

Our results found a significant association between the reduction in the number of reported unhealthy mental health days and member engagement with clinical sessions, but not with coaching sessions. Coaching, and even more so text-based coaching, differs fundamentally in their objectives and practices compared to clinical care [27,28]. Little is understood regarding the effects of text-based coaching on mental health outcomes. Our findings suggest that the amount of care needed to drive member improvement might vary between text-based coaching and clinical practices [29]. It is possible that additional time/sessions might be needed for coaching goals to be formed, implemented, and subsequently have an impact on behavioral change via a text-based medium [28-30]. Future studies should extend the follow-up window when evaluating coaching sessions and assess alternate trajectories of improvement in mental health (eg, nonlinear).

**Limitations**

There are several limitations to consider. One limitation is the potential for bias in our estimates and the increased likelihood that our results may not generalize to all individuals who engage with teletherapy. Furthermore, our cohort design did not have
a comparison group or random assignment to the treatment intervention. Thus, our ability to draw causal inferences is limited and improvements in reported unhealthy days could simply be due to a passage of time; however, we were able to demonstrate significant changes in members with both subclinical and clinical symptoms using real-world longitudinal data. Even though data were missing at random and may not bias results, future studies should implement procedures (eg, incentives) to encourage and capture more complete follow-up data. Future studies can also examine obstacles and facilitators for engagement in teletherapy. The study was also limited to available self-reported outcome data, and there was a large amount of attrition in members reporting unhealthy days over time. This could be due to most members not experiencing clinically meaningful baseline symptomatology and potentially quick improvements in functioning. It is also possible that because the survey was administered outside of the Ginger platform (ie, Survey Monkey), the additional step of completing the measure might have been an added time burden. However, this approach allowed us to pilot and demonstrate the real-world attrition rate when using external data collection platforms.

Conclusions

To our knowledge, this study is one of the first to use the HRQoL measure as a primary outcome in an evaluation of a digital behavioral health platform. Using real-world longitudinal data, our preliminary yet promising results show that short-term engagement with virtual care can be an effective means to improve HRQoL for members with subclinical and clinical symptoms. Virtual care represents a scalable and well-suited approach to meet the growing need for mental health services that has outpaced the in-person availability of clinical mental health professionals. Future studies should examine the long-term impact of text-based coaching and clinical support on HRQoL.

Acknowledgments

We are grateful for the participation of the members who completed the survey, as well as the hard work and dedication of Ginger coaches and clinicians.

Conflicts of Interest

All authors are paid employees of Ginger.

References


Abbreviations

CDC: Centers for Disease Control and Prevention
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition)
GAD-7: General Anxiety Disorder-7
HRQoL: health-related quality of life
PHQ-9: Patient Health Questionnaire-9
Original Paper

Comparison of the Impact of Insulin Degludec U100 and Insulin Glargine U300 on Glycemic Variability and Oxidative Stress in Insulin-Naive Patients With Type 2 Diabetes Mellitus: Pilot Study for a Randomized Trial

Pavle Vrebalov Cindro¹, MD; Mladen Krnić²,3, MD, PhD; Darko Modun¹, MD, PhD; Jonatan Vuković¹, MD, PhD; Tina Tičinović Kurir²,3, MD, PhD; Goran Kardum⁴, MD, PhD; Doris Rušić⁴, PharmD; Ana Šešelja Perišin⁴, PharmD; Josip Bukić⁴, PharmD

¹Department of Gastroenterology, University Hospital Split, Split, Croatia
²Department of Endocrinology, University Hospital Split, Split, Croatia
³Department of Pathophysiology, School of Medicine, University of Split, Split, Croatia
⁴Department of Pharmacy, School of Medicine, University of Split, Split, Croatia
⁵Department of Psychology, Faculty of Humanities and Social Sciences, University of Split, Split, Croatia

Corresponding Author:
Mladen Krnić, MD, PhD
Department of Endocrinology
University Hospital Split
Šoltanska 1
21000 Split
Croatia
Phone: 385 981772113
Email: mladen.krnic@gmail.com

Abstract

Background: There is an ongoing discussion about possible differences between insulin degludec (IDeg-100) and glargine U300 (IGlar-300). There is little data and head-to-head comparison of IDeg-100 and IGlar-300 regarding their simultaneous impact on glycemic variability and oxidative stress in patients with type 2 diabetes mellitus (T2DM).

Objective: In our randomized, open-label, crossover study, we compared the impact of IDeg-100 and IGlar-300 on glycemic variability and oxidative stress in insulin-naive patients with T2DM.

Methods: We recruited a total of 25 adult patients with T2DM (7 females) whose diabetes was uncontrolled (HbA₁c ≥7.5%) on two or more oral glucose-lowering drugs; a total of 22 completed the study. Mean age was 57.3 (SD 6.99) years and duration of diabetes was 9.94 (SD 5.01) years. After the washout period, they were randomized alternately to first receive either IDeg-100 or IGlar-300 along with metformin. Each insulin was administered for 12 weeks and then switched. At the beginning and end of each phase, biochemical and oxidative stress parameters were analyzed. On 3 consecutive days prior to each control point, patients performed a 7-point self-monitoring of blood glucose profile. Oxidative stress was assessed by measuring thiol groups and hydroperoxides (determination of reactive oxygen metabolites test) in serum.

Results: IGlar-300 reduced mean glucose by 0.02-0.13 mmol/L, and IDeg-100 reduced glucose by 0.10-0.16 mmol/L, with no significant difference. The reduction of the coefficient of glucose variation also did not show a statistically significant difference. IGlar-300 increased thiols by 0.08 µmol/L and IDeg-100 increased thiols by 0.15 µmol/L, with no significant difference (P=.07) between them. IGlar-300 reduced hydroperoxides by 0.040 CARR U and IDeg-100 increased hydroperoxides by 0.034 CARR U, but the difference was not significant (P=.12).

Conclusions: The results of our study do not show a significant difference regarding glycemic variability between patients receiving either insulin IDeg-100 or IGlar-300, although IGlar-300 showed greater dispersion of data. No significant difference in oxidative stress was observed. In a larger study, doses of insulins should be higher to achieve significant impact on glycemic parameters and consequently on glycemic variability and oxidative stress.

Trial Registration: ClinicalTrials.gov, NCT04692415; https://clinicaltrials.gov/ct2/show/NCT04692415
Introduction

Background
Global diabetes prevalence in 2019 is estimated to be 9.3% (463 million people), rising to 10.2% (578 million) by 2030 and 10.9% (700 million) by 2045 [1].

The main feature of diabetes mellitus of all types is dysglycemia, which consists of two main components: chronic sustained hyperglycemia and acute glycemic fluctuations from peaks to nadirs. Although disputed by some authors [2], it is generally considered that both components contribute to diabetes complications through two main mechanisms—excessive protein glycation and activation of oxidative stress [3]—with glycemic variability being more specific in having an effect on oxidative stress than chronic sustained hyperglycemia [4], as both upward (postprandial glucose increments) and downward (interprandial glucose decrements) changes activate the oxidative stress pathway [5].

Glucose fluctuations gradually increase from normal glucose metabolism to impaired glucose regulation and diabetes mellitus. Intraday glucose variability occurs at the early stage of abnormal glucose tolerance. In addition to elevated intraday glucose fluctuations, newly diagnosed, drug-naive patients with type 2 diabetes mellitus (T2DM) also demonstrate increased postprandial glucose excursions, higher glucose levels overnight, and more interday fluctuations [6].

The main purpose of insulin therapy in diabetes mellitus is to control glucose—in other words, to combat dysglycemia. Long-acting basal insulin analogues (insulin glargine U100, insulin detemir) significantly improved diabetes management, providing longer duration, flatter profiles of action, lower risk of hypoglycemia, and less glycemic variability compared to NPH (Neutral Protamine Hagedorn) insulin [7,8].

The second generation of basal insulin analogues—insulin degludec 100 units/mL (IDeg-100) and insulin glargine 300 units/mL (IGlar-300)—have even smoother pharmacokinetic/pharmacodynamic profiles than insulin glargine U100, are longer acting, and further lower glycemic variability, at least in patients with T1DM [9,10].

Although several studies [11-13] have compared the impact of these two second-generation basal insulin analogues on glycemic variability in patients with type 1 diabetes mellitus (T1DM), there is little data and head-to-head comparison of IDeg-100 and IGlar-300 regarding their simultaneous impact on glycemic variability and oxidative stress in patients with T2DM. In addition, the results from the T1DM studies are inconsistent [12,13].

Aim of the Study
In this initial study, we compared the impact of IDeg-100 and IGlar-300 on glucose variability and oxidative stress (represented through its surrogate markers) in insulin-naive patients with T2DM. The main research question was whether there is any difference between the two insulins regarding these parameters. The results of this study should inform a larger study comparing these two insulins.

Methods

Ethics Approval and Consent to Participate
This randomized, open-label, crossover study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the University of Split School of Medicine (number 2181-198-03-04-17-0045). All subjects gave written consent prior to their participation in the study.

Study Protocol and Population
Between December 2018 and May 2019, we recruited 25 outpatient insulin-naive patients with uncontrolled T2DM (glycated hemoglobin [HbA₁c] ≥7.5% on two or more oral glucose-lowering drugs) and assigned them to either degludec insulin or glargine U300 insulin combined with metformin. All patients were recruited and treated at University Hospital Split, Croatia. All patients finished the study, but only 22 were analyzed (3 patients were excluded from data analysis—one patient did not perform his 7-point self-monitoring of blood glucose (SMBG) profile 3 days prior to control points, one patient decided to continue with his oral glucose-lowering agents only, and one patient left the study for personal reasons). Basal characteristics of the participants studied are shown in Table 1. The protocol of the study is shown in Figure 1. The study adheres to CONSORT (Consolidated Standards of Reporting Trials) guidelines (Multimedia Appendix 1 displays the CONSORT flow diagram). Patients and the public were not involved in the design, conduct, reporting, or dissemination plans of our research. The trial was retrospectively registered on ClinicalTrials.gov on December 31, 2020 (NCT04692415).
Table 1. Basal clinical characteristics of patients (N=25).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>56.23 (8.09)</td>
</tr>
<tr>
<td>Duration of diabetes (years)</td>
<td>8.90 (5.05)</td>
</tr>
<tr>
<td>Body weight (kg)</td>
<td>89.02 (13.82)</td>
</tr>
<tr>
<td>Body height (cm)</td>
<td>176.20 (10.03)</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>28.92 (3.89)</td>
</tr>
<tr>
<td>Waist circumference (cm)</td>
<td>102.84 (8.56)</td>
</tr>
<tr>
<td>Glycated hemoglobin (HbA₁c), %</td>
<td>9.66 (1.65)</td>
</tr>
<tr>
<td>Fasting glucose (mmol/L)</td>
<td>13.02 (4.47)</td>
</tr>
<tr>
<td>Serum creatinine (µmol/L)</td>
<td>68.24 (13.15)</td>
</tr>
<tr>
<td>Serum uric acid (µmol/L)</td>
<td>310.04 (60.75)</td>
</tr>
<tr>
<td>Total cholesterol (mmol/L)</td>
<td>5.28 (1.46)</td>
</tr>
</tbody>
</table>

Figure 1. The study protocol.

Patients who were eligible for the study fulfilled all of the following inclusion criteria: history of T2DM for at least 1 year, aged between 45 and 65 years (women obligatory in postmenopause and with no hormonal replacement), uncontrolled glycemia on two or more oral antidiabetic drugs, no prior use of insulin, HbA₁c ≥7.5%, receiving statins (if not on statins, they were put on it), not on antiaggregant therapy (if on antiaggregants, they were temporarily excluded from therapy).

Exclusion criteria were the following: the use of glitazones or anticoagulant therapy, renal impairment with creatinine clearance <60 mL/s, presence of malignant disease, chronic liver disease, severe cardiovascular disease and history of cardiovascular incidents (stroke, myocardial infarction, peripheral amputation), and rheumatic and autoimmune diseases.

All participants were asked to avoid the consumption of vitamin supplements, coffee, wine, and Coca-Cola and similar beverages, especially in the days before each control point. Patients were also asked to avoid intensive physical activity up to two days before each control point. All subjects were told to report any side effects immediately, and were given the telephone numbers of the study conductors for this purpose. At each control point, participants were asked about possible side effects.

At baseline, all patients discontinued their previous therapy and were given metformin alone (2 g/day) for 7 days (washout period). After the 7-day washout period, they were randomized alternately by investigators (1:1 ratio) to first receive either IDeg-100 or IGlar-300 subcutaneously according to the order they were included in the study. In phase one, they received either IDeg-100 or IGlar-300 combined with metformin for 12 weeks.
weeks. Phase one was followed by a second washout period in which patients received metformin alone again for 7 days. Finally, in phase two, which also lasted for 12 weeks, patients were switched from IDeg-100 to IGlarr-300 and vice versa (and metformin was continued). The initial dose of both insulins was 0.2 IU/kg. We did not change the dose of insulin during the study period to avoid hypoglycemia, which could significantly influence the results [14-16].

At the beginning and end of each phase, blood samples were collected for the analysis of standard biochemical and oxidative stress parameters (control points 1-4). On 3 consecutive days prior to each control point (at the beginning and the end of each phase), patients completed the 7-point SMBG profile. All patients were already experienced with the use of SMBG [17].

**Glucose Measurement**

To standardize results, all patients received a standard Bionime GM550 glucose meter. They were asked to regularly check their blood glucose 1-2 times per day during the entire study and, in the 3 consecutive days prior to each control point, to perform the 7-point SMBG profile. The 7-point blood glucose profile consisted of seven measurements: (1) before breakfast, (2) 2 hours after breakfast, (3) before lunch, (4) 2 hours after lunch, (5) before dinner, (6) 2 hours after dinner, and (7) before sleeping. Glucose variability was determined by calculating mean glycemia, SD, and coefficient of variation (CV) for each control point [17,18].

**Standard Laboratory Measurement**

Serum uric acid concentrations, serum creatinine, total cholesterol, serum bilirubin values, and other basic biochemical laboratory values were determined by Olympus AU 600 Chemistry Analyzer (Olympus Nichima Co Ltd) and enzymatic laboratory kit.

**Oxidative Stress Measurement**

Thiol groups were assayed according to the Ellman assay [19], modified by Hu [20]. In detail, 100 µL of plasma was diluted with 2 mL of Tris-EDTA buffer (0.1 mol/L Tris, 1 mmol/L EDTA, pH 8.2), and mixed with 100 µL of 10 mM DTNB (5,5′-dithiobis(2-nitrobenzoic acid)), previously prepared in methanol. To subtract the absorbance of plasma and DTNB at 412 nm, two parallel blank samples were assembled. The first one (“blank sample”) was prepared by mixing 2 mL of Tris-EDTA buffer with 100 µL of plasma and the second one (“blank reagent”) was prepared by mixing 2 mL of Tris-EDTA buffer with 100 µL of DTNB. All measurements were performed in triplicate and blanks were run for each sample. Readings were taken spectrophotometrically (Lambda 25; Perkin Elmer) at 505 nm after incubation. Results were compared with a standard curve prepared daily with different concentrations of H_2O_2. The results are expressed in CARR U (Carratuellini units), where 1 CARR U corresponds to 0.08 mg/100 mL H_2O_2.

**Statistical Analysis**

The number of subjects to include in the protocol was selected according to previously published literature [4,23]. Statistical analyses were performed using Statistica 6.0 (StatSoft Inc). Two-way ANOVA for repeated measures was used to evaluate changes in plasma glucose levels, CV, plasma thiols, and hydroperoxides due to IDeg-100 and IGlarr-300.

**Results**

Out of 25 randomized patients, 3 patients were excluded from data analysis—one patient did not perform his 7-point SMBG profile 3 days prior to control points, one patient decided to continue with his oral glucose-lowering agents only, and one patient left the study for personal reasons. No adverse reactions or unintended effects were noticed.

A total of 22 patients (7 females) successfully completed the trial, and their mean basal values were as follows: age, 57.3 (SD 6.99) years; duration of diabetes, 9.94 (SD 5.01) years; body weight, 88.25 (SD 13.57) kg; body height, 174.95 (SD 9.67) cm; BMI, 29.10 (SD 3.80) kg/m^2; waist circumference, 102.73 (SD 8.02) cm; HbA1c, 9.60% (SD 1.68%); fasting glucose, 13.20 (SD 4.48) mmol/L; serum creatinine, 66.0 (SD 2.09) µmol/L; serum uric acid, 305.32 (SD 62.60) µmol/L; total cholesterol, 6.99 (SD 4.48) mmol/L; total triglycerides, 1.21 (SD 0.29) mmol/L; body fat, 25.98 (SD 7.61) %.

The determination of reactive oxygen metabolites (d-ROM) assay measures the concentration of total hydroperoxides in serum or heparin plasma. The method was first described by Alberti et al in 1999 [21] and modified by Verde et al [22], and this modified assay was used to determine d-ROM values in plasma in this study. Each sample was prepared by mixing 2 mL of 0.1 M sodium acetate buffer (pH 4.8) and 20 µL of 0.1 M DMPD (N,N-diethyl-p-phenylenediamine) with 10 µL of plasma. After preparation, tubes with samples were vortexed for 15 seconds and incubated in a thermodimer (Thermomixer comfort, Eppendorf) at 37 °C and 1000 revolutions per minute for 75 minutes. All measurements were performed in triplicate and blanks were run for each sample. Readings were taken spectrophotometrically (Lambda 25; Perkin Elmer) at 505 nm after incubation. Results were compared with a standard curve prepared daily with different concentrations of H_2O_2. The results are expressed in CARR U (Carratuellini units), where 1 CARR U corresponds to 0.08 mg/100 mL H_2O_2. The number of subjects to include in the protocol was selected according to previously published literature [4,23]. Statistical analyses were performed using Statistica 6.0 (StatSoft Inc). Two-way ANOVA for repeated measures was used to evaluate changes in plasma glucose levels, CV, plasma thiols, and hydroperoxides due to IDeg-100 and IGlarr-300.

**Results**

Out of 25 randomized patients, 3 patients were excluded from data analysis—one patient did not perform his 7-point SMBG profile 3 days prior to control points, one patient decided to continue with his oral glucose-lowering agents only, and one patient left the study for personal reasons. No adverse reactions or unintended effects were noticed.

A total of 22 patients (7 females) successfully completed the trial, and their mean basal values were as follows: age, 57.3 (SD 6.99) years; duration of diabetes, 9.94 (SD 5.01) years; body weight, 88.25 (SD 13.57) kg; body height, 174.95 (SD 9.67) cm; BMI, 29.10 (SD 3.80) kg/m^2; waist circumference, 102.73 (SD 8.02) cm; HbA1c, 9.60% (SD 1.68%); fasting glucose, 13.20 (SD 4.48) mmol/L; serum creatinine, 66.0 (SD 2.09) µmol/L; serum uric acid, 305.32 (SD 62.60) µmol/L; total cholesterol, 6.99 (SD 4.48) mmol/L; total triglycerides, 1.21 (SD 0.29) mmol/L; body fat, 25.98 (SD 7.61) %.

The determination of reactive oxygen metabolites (d-ROM) assay measures the concentration of total hydroperoxides in serum or heparin plasma. The method was first described by Alberti et al in 1999 [21] and modified by Verde et al [22], and this modified assay was used to determine d-ROM values in plasma in this study. Each sample was prepared by mixing 2 mL of 0.1 M sodium acetate buffer (pH 4.8) and 20 µL of 0.1 M DMPD (N,N-diethyl-p-phenylenediamine) with 10 µL of plasma. After preparation, tubes with samples were vortexed for 15 seconds and incubated in a thermodimer (Thermomixer comfort, Eppendorf) at 37 °C and 1000 revolutions per minute for 75 minutes. All measurements were performed in triplicate and blanks were run for each sample. Readings were taken spectrophotometrically (Lambda 25; Perkin Elmer) at 505 nm after incubation. Results were compared with a standard curve prepared daily with different concentrations of H_2O_2. The results are expressed in CARR U (Carratuellini units), where 1 CARR U corresponds to 0.08 mg/100 mL H_2O_2. The number of subjects to include in the protocol was selected according to previously published literature [4,23]. Statistical analyses were performed using Statistica 6.0 (StatSoft Inc). Two-way ANOVA for repeated measures was used to evaluate changes in plasma glucose levels, CV, plasma thiols, and hydroperoxides due to IDeg-100 and IGlarr-300.
Figure 2. Mean glucose change, SD, and CV on the first, second, and third day. CV: coefficient of variation; IDeg-100: insulin degludec; IGlar-300: insulin glargine U300. A) Mean glucose change on the first day. B) SD of glucose change on the first day. C) CV of glucose change on the first day. D) Mean glucose change on the second day. E) SD of glucose change on the second day. F) CV of glucose change on the second day. G) Mean glucose change on the third day. H) SD of glucose change on the third day. I) CV of glucose change on the third day.

The SD of glucose excursions was 0.36 for IGlar-300 and 0.06 for IDeg-100, which was not significant ($P=.20$; 95% CI $-0.27$ to 0.87); in addition, there was no statistically significant difference between the two insulins ($P=.07$; 95% CI $-1.07$ to 1.22) (Figure 2B).

The CV on the first day was 0.37 (37%) for IGlar-300 and 0.21 for IDeg-100, which was statistically insignificant ($P=.22$; 95% CI $-0.30$ to 0.63). When compared, CV for these two insulins was not significantly different ($P=.20$; 95% CI $-0.73$ to 1.15) (Figure 2C).

On the second of the 3 days of 7-point SMBG, performed at the end of the observed period, IGlar-300 and IDeg-100 reduced mean glucose by 0.03 and 0.10 mmol/L, respectively, which was statistically insignificant ($P=.08$; 95% CI $-0.09$ to 0.24); there was also no significant difference between the two insulins ($P=.07$; 95% CI $-0.41$ to 0.26) (Figure 2D).

The SD of glucose excursions on the second day was 0.12 for IGlar-300 and –0.05 for IDeg-100, which was insignificant ($P=.19$; 95% CI $-0.23$ to 0.59); when we compared the SD of the two insulins, there was no significant difference ($P=.17$; 95% CI $-1.00$ to 0.65) (Figure 2E).

The CV for the second day was 0.24 (24%) for IGlar-300 and 0.04 for IDeg-100 and that was statistically insignificant ($P=.20$; 95% CI $-0.24$ to 0.63). When compared, CV for the two insulins was not significantly different ($P=.08$; 95% CI $-0.96$ to 0.78) (Figure 2F).

On the third (last) day of the SMBG, the insulins reduced mean glucose levels by 0.04 (IGlar-300) and 0.11 mmol/L (IDeg-100), which was statistically insignificant ($P=.08$; 95% CI $-0.10$ to 0.24), again without a significant difference between the two insulins ($P=.20$; 95% CI $-0.55$ to 0.14) (Figure 2G).
The SD of glucose concentrations on the third day was 0.13 for IGlar-300 and –0.09 for IDeg-100, which was insignificant ($P=.18$; 95% CI –0.15 to 0.61), and comparison of the SD of the two insulins revealed no statistical difference ($P=.14$; 95% CI –0.62 to 0.91) (Figure 2H).

The CV on the third day was 0.19 for IGlar-300 and –0.01 for IDeg-100, which was statistically insignificant ($P=.16$; 95% CI –0.14 to 0.55). When compared, CV for these two insulins was not statistically different ($P=.54$; 95% CI –1.23 to 0.37) (Figure 2I).

IGlar-300 increased thiol levels by 0.08 µmol/L and IDeg-100 increased thiol levels by 0.15 µmol/L ($P=.07$; 95% CI –0.21 to 0.08). No significant difference was found between the two insulins regarding the increase of thiols ($P=.14$; 95% CI –0.15 to 0.44) (Figure 3A).

Although IGlar-300 decreased hydroperoxides by 0.04 CARR U, and IDeg-100 increased hydroperoxides by 0.034 CARR U ($P=.06$; 95% CI –0.19 to 0.05), this impact was not statistically significant, and there was no significant difference between the two insulins ($P=.12$; 95% CI –0.13 to 0.37) (Figure 3B).

Discussion

Principal Findings

Since the new generation of insulin analogues, degludec and glargine U300, appeared on the market, there has been an ongoing discussion about the possible advantages of one insulin over the other. The majority of comparisons related to the incidence of hypoglycemia [7,19,21-23], as well as the absorption stability and profile flatness, as possible causes of differences in hypoglycemia tendency were compared [7,13,24].

Absorption stability and profile flatness, if different, should lead to a difference in glycemic variability between the two insulins as variability in insulin absorption represents an important source of glucose variability in these subjects. Variability can relate to the insulin preparation, the injection technique, and the individual [25].

Glycemic variability and the incidence of hypoglycemia are the elements that “upgraded” the diabetological paradigm from the “diabetic triad” (HbA1c, fasting glycemia, and postprandial glycemia) to the “diabetic pentad” (HbA1c, fasting glycemia, postprandial glycemia, hypoglycemia, and glycemic variability) or even “hexad,” if quality of life is included [26].

The final consequences of increased glycemic variability and long-lasting hyperglycemia, as mentioned in the Introduction section, are diabetic complications, both micro- and macrovascular. As glycemic variability contributes to an increase in oxidative stress (one of two main mechanisms leading to the development of diabetic complications) to a greater extent, the need for the comparison of glycemic variability measures and oxidative stress markers in patients exposed to IDeg-100 and IGlar-300 emerges.

Two large, recently published studies (BRIGHT and DELIVER D+), although different in design, found “more similarities than differences” using IDeg-100 and IGlar-300. On the other hand, another large study (CONFIRM) attributed some advantages to IDeg-100. However, these studies, with the exception of the BRIGHT study, focused primarily on hypoglycemia rate and HbA1c outcomes [27-29]. In this study, we wanted to associate glycemic variability with its ultimate consequence—oxidative stress.

Oxidative stress causes the development of diabetic complications through 5 main molecular mechanisms: the polyol pathway, the hexosamine pathway, increased formation of advanced glycation end products, increased expression of the receptors for advanced glycation end products and their activating ligands, and activation of protein kinase C isoforms. In addition, oxidative stress negatively influences the antiatherosclerotic endothelial enzymes: endothelial nitric oxide synthase and prostacyclin synthase. The intracellular reactive oxygen species increase by these mechanisms, then lead to defective angiogenesis in response to ischemia, and activate proinflammatory and epigenetic mechanisms after the normalization of glycemia (“hyperglycemic memory”) [30].

We used a basal-supported oral therapy variant in this study to emphasize the impact of the insulins studied. However, the results of this initial study did not show statistically significant differences both in glycemic variability and in the expression of oxidative stress in our patients. The primary reason for this was that the decrease in glycemic parameters was too small and,
consequently, the impact on glycemic variability and oxidative stress was too weak. The too-small decrease in glycemic parameters was a consequence of using a low dose of insulin. Namely, we administered both insulins at a dose of 0.2 IU per kg of body weight, and we did not titrate the dose for two reasons: to avoid hypoglycemia, which could significantly influence the oxidative stress and glycemic variability results, and to eschew the difference in dosing of two insulins, which could also affect the results. Many studies have shown that hypoglycemia can worsen oxidative stress through, among other mechanisms, a decrease in nitric oxide and a “reperfusion-like” effect [14-16]. If an examinee has experienced hypoglycemia, he or she should be excluded from the study. Nevertheless, in future studies, the dose of insulins administered must be higher. The question is if “treat-to-target” is the best therapy approach, as treating to target inevitably lowers blood sugar toward hypoglycemia. We think that an initial dose of 0.4 IU/kg with very careful titration of the dose will achieve a more desirable glucose level, but still far from the hypoglycemic zone. A smaller-scale titration study to optimize the dose of insulin and the optimal time for oxidative marker readout would be a good in-between step.

Some studies showed a lower incidence of hypoglycemia with IDeg-100 versus IGlar-300, and that also could contribute to the lower levels of variability and oxidative stress observed with degludec [27], although other studies showed no difference between IDeg-100 and IGlar-300 in that regard [28,29]. Previous research has also suggested a somewhat greater potency of IDeg-100, thus titration to target would probably lead to differences in the final doses used and, consequently, make the comparison more difficult [24,28].

A longer exposure period (longer than 12 weeks) would allow for a more expressed impact of each insulin, presumed positive, although some studies showed a negative effect of chronic insulin therapy on oxidative stress [31]. The longer exposure would possibly explain the difference in the simultaneous increase in thiols and hydroperoxides produced by IDeg-100. The increase in thiol group concentrations represents protein oxidative stress reduction and d-ROM gives insight into the acute changes of lipid peroxide oxidation. This should be considered in the context of the negative effect of chronic insulin treatment on oxidative stress, as mentioned above.

We assessed within-day glycemic variability through changes in average glucose levels, SD of glycemic excursions, and the CV, derived from the SD. The 7-point SMBG profile represents the standard method of glucose monitoring, and we used it for the assessment of glycemic variability [17,32]. Diagnostic CGMS (continuous glucose monitoring systems) and the Libre Flash monitoring system would be more precise tools for glycemic variability measurement but, unfortunately, at the time, due to financial reasons, they could not be employed in this initial study. CGMS or the Flash monitoring system would allow the detection of a greater number of daily peaks and nadirs and give a better insight into glycemic variability [33]. Using CGMS, it would be possible to use the mean amplitude of glycemic excursions (MAGE) index as an assessment tool for glycemic variability as well. Hence, in future studies, we highly recommend the use of CGMS.

In this study, we recruited insulin-naive patients who experienced the failure of oral glucose-lowering therapy and needed insulin introduction. We excluded those who were previously on pioglitazone therapy because of the prolonged action of this drug (which was impossible to remove during the 7-day washout period) and its possible influence on the results. Moreover, we standardized the concomitant therapy by introducing the same dose of atorvastatin in all patients and by temporarily removing salicylic acid from the therapy.

### Conclusion

The results of this study do not show a statistically significant difference in glycemic variability between IDeg-100 and IGlar-300. An insufficient dose of insulin was the main reason for the lack of impact on glycemic parameters and, consecutively, on glycemic variability. Probably due to the absence of a difference in glycemic variability, no difference in the oxidative stress level was noticed. A full-scale study should use larger doses of insulins (at least 0.4 IU/kg), and an optimized and adjusted “treat-to-target” algorithm. CGMS should be used instead of the 7-point SMBG profile. The MAGE index derived from the CGMS should be used for the assessment of glycemic variability. Another small titration study could be performed for optimization of the insulin dose and calculation of the sample size for the main study.

### Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors wish to thank Salus d.o.o. Croatia for financial support and donation of glucose meters and test strips. The authors also thank Ms Tanja Vrebalov Cindro for language review.

### Data Availability Statement

The data sets used and/or analyzed in this study are available from the corresponding author on reasonable request.

### Authors’ Contributions

MK contributed to study concept and design. PVC, MK, and TTK contributed to the acquisition of data. DM, DR, ASP, and JB contributed to laboratory analysis. MK, DM, PVC, and JV contributed to analysis and interpretation of data. MK, DM, and PVC contributed to drafting of the manuscript. MK, PVC, DM, and JV critically revised the manuscript for important intellectual
content. GK, DM, and JV contributed to statistical analysis. MK supervised the study. All authors have read and approved the final manuscript.

Conflicts of Interest
MK received honoraria from Sanofi, NovoNordisk, Eli Lilly, Abbott, MSD, Takeda, Novartis, Boehringer Ingelheim, Servier, Lifescan, and AstraZeneca as a speaker and for attendance at advisory boards. MK author has no conflicts of interest associated with this research.
TTK received honoraria from Sanofi, NovoNordisk, Eli Lilly, Abbott, MSD, Takeda, Novartis, Boehringer Ingelheim, Servier, Lifescan, and AstraZeneca as a speaker and for attendance at advisory boards. TTK author has no conflicts of interest associated with this research.
PVC received honoraria from Abbott, Teva, and Takeda as a speaker and a case presenter. PVC author has no conflicts of interest associated with this research.
JV, DM, GK, DR, ASP, and JB declare no conflicts of interest associated with this research.

Multimedia Appendix 1
CONSORT flow diagram.
[PDF File (Adobe PDF File), 420 KB - formative_v6i7e35655_app1.pdf ]

Multimedia Appendix 2
CONSORT 2010 checklist.
[PDF File (Adobe PDF File), 145 KB - formative_v6i7e35655_app2.pdf ]

References
9. Becker RHA, Dahmen R, Bergmann K, Lehmann A, Jax T, Heise T. New insulin glargine 300 Units · mL-1 provides a more even activity profile and prolonged glycemic control at steady state compared with insulin glargine 100 Units · mL-1. Diabetes Care 2015 Apr;38(4):637-643. [ doi: 10.2337/dc14-0006 ] [ Medline: 25150159 ]


**Abbreviations**

- **CGMS**: continuous glucose monitoring system
- **CV**: coefficient of variation
- **d-ROM**: determination of reactive oxygen metabolites
- **HbA₁c**: glycated hemoglobin
- **IDeg-100**: insulin degludec 100 IU/mL
- **IGlar-300**: insulin glargine 300 IU/mL
- **MAGE**: mean amplitude of glucose excursions
- **NPH**: Neutral Protamine Hagedorn
- **SMBG**: self-monitoring of blood glucose
- **T1DM**: type 1 diabetes mellitus
- **T2DM**: type 2 diabetes mellitus
The Association Between COVID-19 Information Sources and Stigma Against Health Care Workers Among College Students: Cross-sectional, Observational Study

Miharu Nakanishi¹,², PhD; Mai Sakai¹, PhD; Gen Takagi³, PhD; Keita Toshi¹, MSc; Koubun Wakashima⁴, PhD; Hatsumi Yoshii¹, PhD

¹Department of Psychiatric Nursing, Tohoku University Graduate School of Medicine, Sendai-shi, Japan
²Research Center for Social Science & Medicine, Tokyo Metropolitan Institute of Medical Science, Setagaya-ku, Japan
³Department of Welfare Psychology, Faculty of General Welfare, Tohoku Fukushi University, Sendai-shi, Japan
⁴Department of Clinical Psychology & Family Psychology, Graduate School of Education, Tohoku University, Sendai-shi, Japan

Corresponding Author: Miharu Nakanishi, PhD
Department of Psychiatric Nursing
Tohoku University Graduate School of Medicine
2-1 Seiryo-machi
Aoba-ku
Sendai-shi, 980-8575
Japan
Phone: 81 227178179
Email: nakanishi-mh@tohoku.ac.jp

Abstract

Background: The COVID-19 pandemic has triggered stigmatic attitudes against health care workers. Some forms of social media may play a role in disseminating stigmatizing messages.

Objective: We aimed to investigate the association between COVID-19 information sources and stigma against health care workers among college students during the pandemic.

Methods: A cross-sectional, observational study was conducted using a web-based platform in the Tohoku region of Japan. College students aged ≥20 years were asked to complete the questionnaire between August 18 and October 31, 2020. Stigma against health care workers was evaluated using a modified Japanese version of the Social Distance Scale. Participants were also asked to rate their perceived vulnerability to infection using the Japanese version of the Perceived Vulnerability to Disease scale.

Results: A total of 281 students from 8 colleges completed the web-based survey. There were 139 (49.5%) participants who used Twitter, 187 (66.5%) who used news websites, and 46 (16.4%) who used the websites of public health agencies as COVID-19 information sources. After adjusting for age, sex, department, and Perceived Vulnerability to Disease scores, the level of stigma did not differ between students who used Twitter and those who did not. Students who used the websites of public health agencies showed a significantly less stigmatic attitude than those who did not.

Conclusions: Fact-checking and directing visitors to credible information sources from public health agencies may have prevented the formation of stigmatic attitudes toward health care workers. An effective strategy to enable easy access to information provided by public agencies should be integrated into widespread web-based platforms.

(JMIR Form Res 2022;6(7):e35806) doi:10.2196/35806

KEYWORDS
health personnel; social media; social stigma; young adult; Twitter; public health; COVID-19; health care workers; students; information accuracy; information credibility; dissemination; information source; misinformation; information spread; infodemiology
**Introduction**

The COVID-19 pandemic is a serious threat to public health. Moreover, it is characterized by widespread fear, worry, and uncertainty, as many COVID-19 infections are contracted through presymptomatic and asymptomatic transmissions [1]. This health emergency has triggered discriminatory behavior and stigma against health care workers, despite their vital role in caring for people with COVID-19 [2,3]. Stigma is defined as an undesirable characteristic that results in discrimination against an individual [4]. Several incidents of stigmatization of health care workers have been reported; these include avoidance by family members or their community, being denied access to public transport, and even being subjected to physical assault [3,5]. The psychological challenges entailing stigmatization may amplify the negative consequences, such as emotional burnout [6], posttraumatic stress disorder [7], and turnover [8] of working with COVID-19 patients as frontline care providers. A reduction of stigma against health care workers is thus warranted during the COVID-19 pandemic.

The use of unreliable forms of social media as information sources for COVID-19 can lead to the spread of misinformation and increased stigma against health care workers. This is especially true because various COVID-19–related rumors, stigma, and conspiracy theories have been circulating on the internet [9]. Most people rely on the internet for COVID-19 information [10,11]. Particularly, Twitter conversations pertaining to COVID-19 are characterized by the dissemination of stigmatizing messages [12,13]. Although the media, including newspapers, television, and websites, are important sources that can be used to promote health education and literacy, mass media and even health agencies have contributed to the spread of health misinformation that could circulate stigma on the internet [14]. In addition to the media, Twitter allows users to post short messages (tweets), “retweet” messages (reposts), send replies, and “like” messages by other users. Therefore, Twitter users are more likely to be exposed to misinformation and stigmatizing messages, which in turn may exacerbate fear and anxiety and result in the further circulation of such messages [9]. Adolescents and young adults have been using social networking services more frequently since the COVID-19 pandemic, as pandemic-related restrictions have substantially changed their social lives due to school closures [15]. A previous study indicated that the level of anxiety about COVID-19 differed according to the type of preferred news source [16]. However, no studies have examined the association between the types of COVID-19 information sources and stigma against health care workers. An understanding of this association will provide a basis for developing strategies aimed at reducing stigma.

This study, thus, investigated the association between COVID-19 information sources and stigma against health care workers among college students during the pandemic. We hypothesized that college students who used Twitter as a COVID-19 information source would show a more stigmatic attitude toward health care workers than those who do not use Twitter.

**Methods**

**Study Design**

This cross-sectional, observational study was conducted using Google Forms, which is a web-based tool that allows data collection through personalized surveys. An anonymous questionnaire was uploaded and shared through an invitation email to potential participants.

**Setting**

On August 18, 2020, the survey link was shared with teachers from 8 colleges in the Tohoku region of Japan. The link provided a detailed explanation of the study purpose instructions. Subsequently, the teachers emailed the invitation link with the explanatory documents to the students. Participants aged ≥20 years were requested to complete the questionnaire on October 31, 2020.

The first page of the website contained details on the voluntary nature of the participation and protection of personal information. After reading the introduction, students indicated their consent to participate by clicking on the link to start the survey. The “Limit to 1 response” function of Google Forms was enabled to prevent respondents from completing the form more than once.

**Participants**

We used purposive and convenience sampling to select colleges and departments with (1) nursing students who underwent on-the-job training in hospitals and (2) students from other departments. The other departments were selected to ensure gender ratios similar to those of students in the nursing departments.

Nursing and other students were recruited to compare the level of stigma against health care workers. We assumed that nursing students would show less stigmatic attitudes than other students, because they are expected to become professional nurses and, therefore, have more psychological proximity with other health care workers.

**Measurements**

The questionnaire included questions regarding stigmatic attitudes against health care workers, COVID-19 information sources, perceived vulnerability to infection, department, age, sex, and contact with COVID-19 patients. The questions and response options of stigma against health care workers were developed for this study (Table S1 in Multimedia Appendix 1).

Stigma against health care workers was evaluated using a modified version of the Japanese version of the Social Distance Scale (SDSJ) [17]. The original Social Distance Scale was developed based on the Keyed Favorable Response and Ego-Involvement Ratings of Scale [18], which is used to assess the level of stigmatic attitude toward patients with schizophrenia. It contains 8 stigma-related items, which are rated on a 4-point Likert scale, and the scale has good reliability and validity [17]. For this study, we replaced “patients with schizophrenia” in each item with “health care workers and their families who are performing infectious disease management with an unestablished
The total score on the 8 items ranged from 0 to 24; higher scores indicated more stigmatic attitudes.

The 8 types of COVID-19 information sources were listed in the questionnaire. Participants were asked to check any source, if applicable. The list was developed by a research panel to contain a range of social media platforms, including newspapers, television news streams, television tabloid talk shows, news websites, Twitter, the websites of public health agencies (e.g., Ministry of Health, Labour and Welfare), Instagram, and Facebook.

Perceived vulnerability to infection was evaluated using the Perceived Vulnerability to Disease (PVD) scale [19]. The PVD contains 15 items, which are rated on a 7-point Likert scale from 1 to 7. It comprises 2 subscales: Perceived Infectability and Germ Aversion. Perceived Infectability refers to beliefs about immunological functioning and personal susceptibility to infectious diseases. Germ Aversion refers to aversive affective responses to situations that connote a relatively high likelihood of pathogen transmission. The Japanese version of the PVD is reported to have good reliability and validity [20]. In this study, the Cronbach’s coefficient was .78 (95% CI .74-.82) for Perceived Infectability and .65 (95% CI .58-.71) for Germ Aversion. Although Perceived Infectability refers to one’s susceptibility to infection, Germ Aversion covers behaviors exerting emotional discomfort in a high-pathogen context, which in turn deters from the source of infection. Therefore, we assumed that stigma against health care workers would show a moderate positive correlation with Germ Aversion but not with Perceived Infectability.

Participants also answered items pertaining to their age, sex, department, and contact with COVID-19 patients. The presence of contact was assessed if (1) the respondent or their family members or friends had been infected with COVID-19; and if (2) the respondent or their family members or friends had close contact with infected persons. Based on the contact hypothesis [21,22], we assumed that individuals who directly interacted with COVID-19 patients would be less likely to hold stigmatic attitudes against health care workers. However, the number of individuals with direct contact was small (n=18). Therefore, we used the presence of contact for sensitivity analysis instead of including it in bivariate and multivariate analyses.

Study Size
The required sample size was calculated using G*Power (version 3.1.9.7; Faul et al [23,24]). Based on a recent report on the use of Twitter among Japanese college students [25], we assumed the prevalence of Twitter as a COVID-19 information source to be 50% in this study. Assuming an α level of 5%, 95% power, and medium effect size (Cohen d=0.5) for a 2-tailed test, the minimum sample size was determined to be 210.

Statistical Analysis
The validity and reliability of the modified SDSJ were examined. To test concurrent validity, the mean difference between nursing and other students was examined. The normality of distribution was assumed for total SDSJ score (Table S2 in Multimedia Appendix 1). The equality of variances between the 2 groups was also assumed (Table S3 in Multimedia Appendix 1). Therefore, 2-sided testing using 2-tailed students’ t test was used. To test convergent validity, Pearson correlation coefficient was calculated between the total SDSJ score and the subscales scores of the PVD. To test internal reliability, the Cronbach α coefficient and 95% CI [26] were calculated for the total SDSJ score.

Within the types of COVID-19 information sources, overlaps between Twitter and other major web-based platforms (news websites and the websites of public health agencies) were examined. To investigate the association between Twitter use and stigma against health care workers, the mean difference in SDSJ scores was examined by performing a 2-tailed t tests between participants who used Twitter as a COVID-19 information source and those who did not. Furthermore, a multiple Ordinary Least Square (OLS) regression analysis was performed using the SDSJ scores as the dependent variable and the types of information sources as independent variables. Participant characteristics, including age, sex, department, and PVD scores, were included as covariates. Diagnostic tests were conducted to test the assumptions of OLS regression, including the normality of the residuals, homoskedasticity, the absence of outliers, and low multicollinearity (Table S4 in Multimedia Appendix 1). Since some potential outliers were found, a sensitivity analysis of the multivariate model was performed by excluding them. Another sensitivity analysis was also conducted by excluding individuals with social contacts.

All analyses were conducted using Stata statistical software (version 17.0; StataCorp). The significance level was set at low (α=.1), medium (α=.05), and high (α=.01) [27]. As our primary endpoint was to assess the association between use of Twitter and SDSJ under multiple regression analysis adjusting for covariates, we did not apply P value adjustments for multiple hypothesis testing [28].

Ethics Approval
The study protocol was approved by the Ethics Board of Tohoku University (2021-1-733) and was conducted in accordance with the Helsinki Declaration of 1975 (as revised in 2013).

Results
Participant Characteristics
A total of 281 participants completed the survey and were included in the final sample. The majority (86.7%, n=238) were women and the most frequent (50.9%, n=143) age was 21 years. There were 18 participants who reported coming into contact with COVID-19 (Table 1).
Table 1. Participant characteristics.

<table>
<thead>
<tr>
<th>Variable, category</th>
<th>Participant (N=281), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>238 (86.7)</td>
</tr>
<tr>
<td><strong>Age (year)</strong></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>86 (30.6)</td>
</tr>
<tr>
<td>21</td>
<td>143 (50.9)</td>
</tr>
<tr>
<td>≥22</td>
<td>52 (18.5)</td>
</tr>
<tr>
<td><strong>Department</strong></td>
<td></td>
</tr>
<tr>
<td>Nursing</td>
<td>104 (37)</td>
</tr>
<tr>
<td>Rehabilitation</td>
<td>70 (24.9)</td>
</tr>
<tr>
<td>Psychology</td>
<td>63 (22.4)</td>
</tr>
<tr>
<td>Other</td>
<td>44 (15.7)</td>
</tr>
<tr>
<td><strong>Contact with COVID-19</strong></td>
<td></td>
</tr>
<tr>
<td>Any of the experiences below</td>
<td>18 (6.4)</td>
</tr>
<tr>
<td>Family members or friends had close contact with an infected person</td>
<td>13 (4.6)</td>
</tr>
<tr>
<td>I had close contact with an infected person</td>
<td>1 (0.4)</td>
</tr>
<tr>
<td>Family members or friends had been infected with COVID-19</td>
<td>5 (1.8)</td>
</tr>
<tr>
<td>I had been infected with COVID-19</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

**Stigma Against Health Care Workers**

The mean total SDSJ score in overall sample was 7.9 (SD 4.7). Nursing students had a significantly lower mean total score (mean 6.0, SD 4.5) than other students (mean 8.9, SD 4.5; $t_{218.48}=5.18$, $P<.001$). The mean total SDSJ score did not differ between participants with social contact (mean 8.6, SD 1.0) and those without (mean 7.8, SD 0.3; $t_{20.16}=0.73$, $P=.47$).

Pearson correlation coefficients were −0.03 ($P=.62$) between the SDSJ and PVD Perceived Vulnerability scores and 0.33 ($P<.001$) between the SDSJ and PVD Germ Aversion scores. The Cronbach $\alpha$ coefficient of the SDSJ score was high ($\alpha=.83$; 95% CI .80-.86). In summary, the modified SDSJ demonstrated satisfactory concurrent, convergent, and internal validity.

**COVID-19 Information Sources**

Half (49.5%, 139/281) of the participants used Twitter as a COVID-19 information source (Table 2). Since less than 10% of participants used newspapers (8.5%, n=24), Instagram (3.2%, n=9), or Facebook (0.4%, n=1), we excluded these sources from the following multivariate analysis.

In total, 89 participants reported using both Twitter and news websites, of which 16 participants also used the websites of public health agencies. Further, 44 participants used Twitter, but not news websites or the websites of public health agencies (Figure 1).

The total SDSJ score did not differ between Twitter users (mean 8.1, SD 4.4) and those who did not use Twitter (mean 7.6, SD 5.0; $t_{276.63}=0.97$, $P=.33$).
Table 2. Prevalence of COVID-19 information sources.

<table>
<thead>
<tr>
<th>Type, category</th>
<th>Participant (N=281), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newspaper</td>
<td>24 (8.5)</td>
</tr>
<tr>
<td>Television</td>
<td></td>
</tr>
<tr>
<td>News stream</td>
<td>218 (77.6)</td>
</tr>
<tr>
<td>Tabloid talk show</td>
<td>71 (25.3)</td>
</tr>
<tr>
<td>Web-based source</td>
<td></td>
</tr>
<tr>
<td>News website</td>
<td>187 (66.5)</td>
</tr>
<tr>
<td>Twitter</td>
<td>139 (49.5)</td>
</tr>
<tr>
<td>Website of public health agencies (eg, Ministry of Health, Labour and Welfare)</td>
<td>46 (16.4)</td>
</tr>
<tr>
<td>Instagram</td>
<td>9 (3.2)</td>
</tr>
<tr>
<td>Facebook</td>
<td>1 (0.4)</td>
</tr>
<tr>
<td>Other social networking services</td>
<td>13 (4.6)</td>
</tr>
</tbody>
</table>

Figure 1. Venn diagram of overlaps between Twitter and other major web-based platforms (news websites and the websites of public health agencies).

Association Between Stigma and Information Sources
A multiple regression analysis showed that the total SDSJ scores were lower among participants using the websites of public health agencies ($P=.008$), nursing students ($P<.001$), and those with lower Germ Aversion scores ($P<.001$; Table 3). The use of Twitter was not associated with SDSJ scores ($P=.58$).

The results of OLS diagnostic tests showed that the following assumptions were met: the normality of the residuals, homoskedasticity, and low multicollinearity (Table S4 in Multimedia Appendix 1). However, there were 2 potential outliers (Table S4 in Multimedia Appendix 1). A sensitivity analysis that excluded the 2 individuals did not change the results (Table S5 in Multimedia Appendix 1).

Another sensitivity analysis that excluded individuals (n=18) who came into close contact with a patient with COVID-19 did not alter the results (Table S5 in Multimedia Appendix 1).
Table 3. Multiple linear regression analysis of stigma against health care workers.\(^a,b\)

<table>
<thead>
<tr>
<th>Variable, category</th>
<th>Coefficient (95% CI)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>COVID-19 information source</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Television news stream</td>
<td>0.86 (–0.40 to 2.12)</td>
<td>.18</td>
</tr>
<tr>
<td>Television tabloid show</td>
<td>0.51 (–0.70 to 1.72)</td>
<td>.41</td>
</tr>
<tr>
<td>News website</td>
<td>0.60 (–0.45 to 1.66)</td>
<td>.26</td>
</tr>
<tr>
<td>Twitter</td>
<td>0.29 (–0.74 to 1.31)</td>
<td>.58</td>
</tr>
<tr>
<td>Websites of public health agencies</td>
<td>–1.84 (–3.20 to –0.49)</td>
<td>.008</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.07 (–0.38 to 2.52)</td>
<td>.15</td>
</tr>
<tr>
<td><strong>Age (year)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.33 (–1.21 to 1.88)</td>
<td>.67</td>
</tr>
<tr>
<td>21</td>
<td>–0.25 (–1.64 to 1.14)</td>
<td>.72</td>
</tr>
<tr>
<td>≥22</td>
<td>reference</td>
<td>reference</td>
</tr>
<tr>
<td><strong>Department</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nursing</td>
<td>–3.04 (–4.11 to –1.98)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Perceived vulnerability to infection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Infectability</td>
<td>–0.13 (–0.63 to 0.37)</td>
<td>.61</td>
</tr>
<tr>
<td>Germ Aversion</td>
<td>1.85 (1.31-2.39)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

\(^a\)Stigma against health care workers was evaluated using the modified Japanese language version of the Social Distance Scale; the total score ranges from 0 to 24.

\(^b\)Perceived vulnerability to infection was evaluated using the Japanese version of the Perceived Vulnerability to Disease scale; total scores range from 1 to 7 for both Perceived Infectability and Germ Aversion.

**Discussion**

With increasing concerns about the stigma against health care workers during the COVID-19 pandemic, strategies to reduce stigma are the need of the hour, considering the prevalence of Twitter and other information sources for information related to COVID-19. Thus, this study examined the association between the types of information source and stigma against health care workers among college students.

**Principal Findings**

Contrary to our hypothesis, Twitter use was not associated with the stigma against health care workers. The survey was conducted between August and October 2020 when the daily number of new COVID-19 cases in Japan ranged from 219 to 1178. This period is a few months after the onset of the outbreak, and our participants may have had a lower level of fear than that from February to April 2020 [9,12,13]. In the acute phase of the psychological response to a crisis, such as the Great East Japan Earthquake on March 11, 2011, Twitter messages diffused rumors and misinformation [29]. However, the level of anxiety expressed in Twitter messages appears to return to normal over time [30]. This shift to Twitter for content has also been observed for COVID-19 [31]. In addition, several students aged ≥20 years in the Tohoku region might have experienced the March 11 earthquake when they were children, resulting in their learned mindset to treat Twitter messages with caution.

College students who used the websites of public health agencies as information sources reported significantly less stigmatic attitudes than those who did not \((P=.008)\) in the multiple regression analysis. Fact-checking and directing users to credible information sources from the websites of public health agencies can prevent the further spread of misinformation [32]. News websites also disseminate official announcements from public health agencies. In this study, substantial overlaps were observed between the users of the websites of public health agencies, news websites, and Twitter. This finding is consistent with a previous study [10]. Furthermore, public health agencies also have Twitter accounts, and each message is limited to 140 Japanese characters. Their messages usually include a URL to official website pages that contain longer texts. The accuracy of the information held by Twitter users may vary between those who only read the message and those who click on the link to access the website. Adequate communication strategies should be embedded in reliable information from trusted sources [31,33]. Considering that Twitter was a popular COVID-19 information source, access to information curated by public agencies may help reduce exposure to misinformation and stigma against health care workers. In addition to accurate information, “hero” messaging [34] should be reserved for health care workers in public policy. Health care workers were deemed essential frontline heroes during the COVID-19 crisis. Such perceptions can mitigate stigmatic attitudes toward health care workers. Furthermore, positive video messages embedded in tweets may enable young people to engage in parasocial interactions with health care workers, which in turn may help...
change their negative beliefs about health care workers based on the parasocial contact hypothesis [22,35].

Stigma against health care workers, as measured by the modified SDSJ, was significantly lower among nursing students (P<.001) and students with lower germ aversion (P<.001). Unlike Germ Aversion, we found very little statistical evidence that Perceived Infectability was associated with stigma against health care workers using the total SDSJ score. These associations were consistent with our assumption that the modified scale is valid.

Strengths and Limitations
The strength of this study lies in the examination of stigma against health care workers in the context of COVID-19 and social media, thus addressing a gap in the literature. A limitation was the lack of information accuracy and limited data on the intensity of social media use among participants. In addition, the cross-sectional design precludes causal inferences between stigma and the types of information sources. The use of websites of public health agencies might indicate the high health literacy and low stigmatic attitudes of students.

Conclusions
A few months after the onset of the COVID-19 pandemic, nearly half of the college student population used Twitter as an information source. Our findings showed that the level of stigma against health care workers did not differ according to Twitter use. Students who used the websites of public health agencies reported less stigmatic attitudes than those who did not. These results imply that directing people to credible COVID-19 information sources from public agencies may prevent the formation of stigmatic attitudes against health care workers. An effective strategy for the induction of access to credible information sources should be explored for integration into Twitter and other widespread web-based platforms.

Acknowledgments
This study was supported by the Department of Psychiatric Nursing, Tohoku University Graduate School of Medicine. The funder was not involved in the design or study course; data collection, management, analysis, or interpretation; or monitoring of the manuscript for presentation, review, or approval.

Authors’ Contributions
MS, KW, and HY collected the data. MN, MS, GT, and KT analyzed and interpreted the data and drafted the manuscript. KW and HY were involved in the study design and setup, the supervision of data analysis, and finalizing the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary tables on (1) questions and response options used in the online survey; (2) test of the normality of scores; (3) Ordinary Least Square regression diagnostic tests; (4) sensitivity analysis: multiple linear regression analysis of stigma against health care workers, excluding possible outliers (N=279); and (5) sensitivity analysis: multiple linear regression analysis of stigma against health care workers, excluding individuals who came in close contact with a patient with COVID-19 (N=263).

References


**Abbreviations**

- **OLS**: Ordinary Least Square
- **PVD**: Perceived Vulnerability to Disease
- **SDSJ**: Japanese version of the Social Distance Scale

©Miharu Nakanishi, Mai Sakai, Gen Takagi, Keita Toshi, Koubun Wakahama, Hatsumi Yoshii. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.07.2022. 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.
Corrigenda and Addenda

Correction: The Associations Between Racially/Ethnically Stratified COVID-19 Tweets and COVID-19 Cases and Deaths: Cross-sectional Study

Xiaohui Liu1,2, PhD; Bandana Kar3, PhD; Francisco Alejandro Montiel Ishino1, MPH, PhD; Tracy Onega2, MPAS, PhD; Faustine Williams1, MPH, PhD

1National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD, United States
2Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, United States
3National Security Sciences Directorate, Oak Ridge National Lab, Knoxville, TN, United States

Corresponding Author:
Xiaohui Liu, PhD
National Institute on Minority Health and Health Disparities
National Institutes of Health
6707 Democracy Boulevard, Suite 800
Bethesda, MD, 20892
United States
Phone: 1 8016464179
Email: xiaohuiliugis@gmail.com

Related Article:
Correction of: https://formative.jmir.org/2022/5/e30371

In “The Associations Between Racially/Ethnically Stratified COVID-19 Tweets and COVID-19 Cases and Deaths: Cross-sectional Study” (JMIR Form Res 2022 May 30;6(5):e30371. doi: 10.2196/30371), one error was noted.

In the original article author Faustine Williams was incorrectly associated with affiliation 2:

Faustine Williams2, MPH, PhD
2Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, United States.

The correct affiliation should be affiliation 1:

Faustine Williams1, MPH, PhD
1National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD, United States.

The correction will appear in the online version of the paper on the JMIR Publications website on July 6, 2022, 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.
Abstract

Background: Dental care expenses are reported to present higher financial barriers than any other type of health care service in the United States. Social media platforms such as Twitter have become a source of public health communication and surveillance. Previous studies have demonstrated the usefulness of Twitter in exploring public opinion on aspects of dental care. To date, no studies have leveraged Twitter to examine public sentiments regarding dental care affordability in the United States.

Objective: The aim of this study is to understand public perceptions of dental care affordability in the United States on the social media site, Twitter.

Methods: Tweets posted between September 1, 2017, and September 30, 2021, were collected using the Snscrape application. Query terms were selected a priori to represent dentistry and financial aspects associated with dental treatment. Data were analyzed qualitatively using both deductive and inductive approaches. In total, 8% (440/5500) of all included tweets were coded to identify prominent themes and subthemes. The entire sample of included tweets were then independently coded into thematic categories. Quantitative data analyses included geographic distribution of tweets by state, volume analysis of tweets over time, and distribution of tweets by content theme.

Results: A final sample of 5314 tweets were included in the study. Thematic analysis identified the following prominent themes: (1) general sentiments (1614 tweets, 30.4%); (2) delaying or forgoing dental care (1190 tweets, 22.4%); (3) payment strategies (1019 tweets, 19.2%); (4) insurance (767 tweets, 14.4%); and (5) policy statements (724 tweets, 13.6%). Geographic distributions of the tweets established California, Texas, Florida, and New York as the states with the most tweets. Qualitative analysis revealed barriers faced by individuals to accessing dental care, strategies taken to cope with dental pain, and public perceptions on aspects of dental care policy. The volume and thematic trends of the tweets corresponded to relevant societal events, including the COVID-19 pandemic and debates on health care policy resulting from the election of President Joseph R. Biden.

Conclusions: The findings illustrate the real-time sentiment of social media users toward the cost of dental treatment and suggest shortcomings in funding that may be representative of greater systemic failures in the provision of dental care. Thus, this study provides insights for policy makers and dental professionals who strive to increase access to dental care.

(JMIR Form Res 2022;6(7):e36315) doi:10.2196/36315

KEYWORDS
dentistry; oral health; social media; access to care; healthcare reform; COVID-19; dental care; health care service; twitter; public health; health communication; dental treatment; health policy; dental professional; thematic analysis
Introduction

A lack of access to dental care can lead to lower levels of systemic health, quality of life, and economic outcomes [1]. Yet, those who are most in need of dental care are often the least likely to receive it [2]. Low-income, working-age adults report the highest levels of financial barriers to needed dental care [3]. In the United States, the financial barriers to accessing dental care are higher than any other type of health care service [3-5]. The percentage of the population without dental insurance is more than twice that of those who are medically uninsured [6]. Spending on dental care results in a high percentage of out-of-pocket expenses because of a lack of insurance or high insurance deductibles and copayments [7].

While comprehensive dental coverage for children is an essential benefit under the Affordable Care Act, dental coverage for adults remains optional [7]. At the time of this study, 3 states provide no dental coverage, and 12 states provide emergency-only dental services to Medicaid beneficiaries [8]. In a comparison of public dental coverage for older adults in high-income countries, the United States had the shallowest dental coverage for older adults [9]. In an effort to mitigate access inequalities, there was a recent push in the US federal government to provide for Medicare coverage of dental and oral health services. In January of 2021, the Medicare Dental Benefit Act of 2021 (H.R.502 and S.97) was introduced into Congress, and President Biden’s budget-reconciliation package proposed funding for a Medicare dental benefit. In August 2021, the Centers for Medicare and Medicaid Services appointed a chief dental officer to guide it in advancing oral health in Medicare [10]. To best develop a policy to address financial barriers to dental care, perspectives at the individual level are needed.

Social media is increasingly becoming an essential tool for public health communication [11]. Twitter is a free social media service where people communicate their daily thoughts and behaviors in short, 280-character messages called “tweets.” With over 68 million active monthly users in the United States, Twitter offers rich, population-based data for tracking concerns of public health significance [12]. Twitter data emerge from real-world social environments, which encompass a large and diverse range of people, without any prompting from researchers. This contrasts with traditional approaches of public surveillance where responses are elicited in the form of semistructured interviews and web-based surveys with open-ended questions [13]. In addition, Twitter is a compelling data source for public health researchers because of the real-time nature of the content and high level of correlation with user sentiment and consumer confidence indices [14]. These qualities have contributed to a growing number of studies examining the use of Twitter for public health research [15-21], including the investigation of aspects of oral health [22-26]. Several studies have aimed to assess the influence of societal events on Twitter content related to oral health [27,28].

This study aims to explore the sentiments of Twitter users in the United States on dental care affordability in order to summarize trends and perceptions that describe how the cost of dental treatment impacts access to dental care. To date, no studies have leveraged Twitter to examine public sentiments regarding dental care affordability in the United States.

Methods

Ethical Considerations

This infodemiological study used a convergent mixed methods approach [29,30] to analyze publicly available Twitter content related to dental care affordability. This study was submitted to the Institutional Review Boards of New York University (IRB-FY2021-5634) and the University of Washington (STUDY00013725). In alignment with federal regulations regarding the use of publicly available data for research, both institutional reviews determined that the study did not meet the criteria for research involving human subjects and that no further review was required.

Data Collection and Preprocessing

Data were obtained from Twitter, a free social media website created in 2006. Tweets can remain visible to the public or can be made visible only to approved followers, at the discretion of the user. Only publicly available tweets were used in this study, and usernames were removed for privacy protection.

Search terms were generated to identify tweets that discussed financial considerations associated with dental treatment. These search terms were tested and expanded through pilot queries and assessments until they were refined to the following word stems: “dental,” “dentist,” “tooth,” “teeth,” “root canal” AND “expensive,” “pay,” “afford,” and “money.”

Using a newly created account on Twitter, tweets were collected between September 1, 2017, and September 17, 2021, using Snscrape [31-33], an open-source web scraper written in Python (Python Software Foundation). The data collection script included a specific query that consisted of any combination of the search terms. Duplicate tweets, foreign language tweets, and retweets were then excluded. In addition to the tweets, metadata was collected as follows: “url,” “date,” “renderedContent,” “user,” “replyCount,” “retweetCount,” “likeCount,” “quoteCount,” “retweetedTweet,” “quotedTweet,” “mentionedUsers,” “coordinates,” “place,” “hashtags,” and “cashtags.”

Metadata related to the users’ location were used in order to limit the included tweets to those in the United States. Content was excluded from the data set of tweets for any of the following reasons: (1) content was unrelated to dental-treatment needs or experiences; (2) content was determined to be an advertisement; (3) content pertained to purely cosmetic or orthodontic dental procedures; (4) content pertained to veterinary dentistry; (5) content was classified as a joke or sarcasm; and (6) content was in reference to dental policies outside of the United States.

Data Analysis

Data were analyzed qualitatively using both deductive and inductive approaches. Two members of the research team with clinical dental knowledge (SY and LB) co-coded all of the tweets. In total, 8% (440/5500) of all of the included tweets were coded in order to identify prominent themes and subthemes. A final codebook was developed through consensus
among the members of the research team. The themes identified for the codebook included (1) general sentiments; (2) delaying or forgoing dental care; (3) payment strategies; (4) insurance; and (5) policy statements. Using the codebook, the entire sample of the included tweets was independently coded by the aforementioned researchers. In instances where the coders disagreed, they reached a consensus through discussion.

Quantitative data analysis included geographic distribution of tweets by state, volume analysis of tweets over time, and distribution of tweets by content theme. The tweets were mapped to an individual state within the United States, and the volume of tweets over time were depicted using Tableau (Tableau Software Inc), a data visualization software.

**Results**

**Data Collection**

Using the search terms, we collected a total of 18,685 tweets from September 1, 2017, to September 30, 2021. Of these tweets, 2617 (14%) were removed as duplicates, and 10,754 (57.6%) were removed based on preestablished exclusion criteria. A final sample of 5314 (28.4%) tweets (from 4963 unique users) were included in the study (Figure 1). The top 5 pairs of search terms were as follows: dental_afford (520/5314, 9.8% tweets); teeth_pay (461/5314, 8.7% tweets); dentist_pay (436/5314, 8.2% tweets); dental_money (356/5314, 6.7%, tweets); and dental_expensive (353/5314, 6.6% tweets).

![Figure 1. Data collection flow chart. Tweets about dental care affordability.](image)

<table>
<thead>
<tr>
<th>Main category</th>
<th>Overall (n=5314), n (%)</th>
<th>2017 (n=277), n (%)</th>
<th>2018 (n=1227), n (%)</th>
<th>2019 (n=1341), n (%)</th>
<th>2020 (n=1090), n (%)</th>
<th>2021 (n=1379), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>General sentiments</td>
<td>1614 (30.4)</td>
<td>110 (39.2)</td>
<td>399 (32.5)</td>
<td>405 (30.2)</td>
<td>320 (29.4)</td>
<td>380 (27.6)</td>
</tr>
<tr>
<td>Delaying or forgoing care</td>
<td>1190 (22.4)</td>
<td>64 (23.1)</td>
<td>304 (24.8)</td>
<td>286 (21.3)</td>
<td>266 (24.4)</td>
<td>270 (19.6)</td>
</tr>
<tr>
<td>Payment strategies</td>
<td>1019 (19.2)</td>
<td>34 (12.3)</td>
<td>206 (16.8)</td>
<td>298 (22.2)</td>
<td>249 (22.8)</td>
<td>232 (16.8)</td>
</tr>
<tr>
<td>Insurance</td>
<td>767 (14.4)</td>
<td>44 (15.9)</td>
<td>173 (14.1)</td>
<td>187 (13.9)</td>
<td>155 (14.2)</td>
<td>208 (15.1)</td>
</tr>
<tr>
<td>Policy statements</td>
<td>724 (13.6)</td>
<td>25 (9.0)</td>
<td>145 (11.8)</td>
<td>165 (12.3)</td>
<td>100 (9.2)</td>
<td>289 (21.0)</td>
</tr>
</tbody>
</table>

**Thematic Distribution of Tweets**

Thematic analysis identified the following prominent themes (Table 1): (1) general sentiments (1614/5314 tweets, 30.4%); (2) delaying or forgoing dental care (1190/5314 tweets, 22.4%); (3) payment strategies (1019/5314 tweets, 19.2%); (4) insurance (767/5314 tweets, 14.4%); and (5) policy statements (724/5314 tweets, 13.6%). The identified subthemes and illustrative quotes from the data are presented below and in Table 2.
Table 2. Representative tweets for theme or subtheme as described in the codebook.

<table>
<thead>
<tr>
<th>Theme and subtheme</th>
<th>Tweet</th>
</tr>
</thead>
</table>
| Insurance                           | ● “My dental insurance deductible was so high, I couldn't have dental work done last year. And my knee surgery got cancelled as a result. Damn shame. Still can't afford it. [URL]”  
● “And... 19's root canal and crown cost more than the annual dental insurance max. There go $837 of tuition money. I'm really just over this whole century.”  
● “According to most health insurance companies teeth and eyes are luxury items that I must pay more to continue enjoying because they're a cosmetic privilege.” |
| Payment strategies                  | ● “@username @username hey. I know it’s a long shot but I just had to use my rent money to have a very badly infected tooth removed and now I don’t know how I’m doing to pay rent. PLEASE help if you can 🙏 God bless ya’ll doing God’s work”a  
● “Nervous to get my tooth pulled tomorrow but mainly because I’m afraid it’s gonna cost more than what I have left on the credit card I’m using to pay for it…” |
| Delaying or foregoing care          | ● “I canceled my root canal that was supposed to be Monday because my tooth quit hurting but here we are in ridiculous pain again and my husband has no job so no money to spend on it right now 😞”  
● “@username Yes it is I have several teeth I need to get pulled/ worked on but can’t afford it...and also take pain meds that don't seem to help much of anything…”a  
● “@username Do you live near a dental school? When I didn’t have dental insurance, I did an expensive project at a dental school for about 1/5 the cost. It took more time, but it was worth paying less.”a |
| General sentiments                  | ● “I’m positive the dentist doesn’t really find cavities in my teeth they just want to get my money and torture me. How do I brush every day and floss and still get cavities every time”  
● “@username Dental work is horribly expensive. I have had my very last $1700 root canal.”a  
● “@username @username This is why I never want to hear another American make fun of British teeth again. People in this country cannot afford the astronomical prices charged by dentists for dental care. So they just go without dental care at all.”a  
● “i spent a lotta money on my teeth and it was worth it 🙌”  
● “@DeptVetAffairs @SenateGOP @HouseGOP I cannot afford regular dental coverage as a disabled veteran with a $200 a month income.”  
● “My mom doesn’t have insurance for dental care...now, she has an infection in her tooth... now she needs to pay $1504 before seeing the specialist... we don’t have $1504 in the bank... This is why we need Medicare For All!!! I’m angry because of this greedy corruption #Bernie2020”  
● “@username @username When someone has a dental problem serious enough for a root canal, they are in pain and it will be an emergency treatment. But only people with money or dental insurance can get one. Most states’ Medicaid doesn’t cover adult dental care, and if it does, they pull the tooth.”a |

*aUsername was removed to maintain the privacy of the Twitter user.

**General Sentiments**

The most prominent theme included tweets containing general sentiments that expressed that dental care was expensive or not affordable but did not additionally suggest the cost that prohibited the user from accessing needed dental care. The overwhelming majority of these tweets expressed negative sentiments. Tweets in this thematic category were subcoded into the following subthemes: mistrust of dentists, expressions that dental care is expensive for the tweeting individual or a personal acquaintance (eg, a family member), impersonal statements about the affordability of dental care, and positive sentiments about accessing dental care.

The most prevalent subtheme was “expensive dental care,” representing 70.3% (n=1135) of the tweets in this theme.

Individuals frequently expressed displeasure related to the direct cost of dental care and surcharges related to administrative costs.

*The consultation alone was $250... out of pocket at that ...why the dentist gotta be so expensive?? I just want to be beautiful and healthy for the low [URL]. [User #677]*

Tweets about “dentist mistrust” represented the second most prevalent subtheme of general sentiments. Users expressed sentiments of being financially duped or perceptions of dentists prioritizing financial gain over the patient’s health.

*@username @username Another common practice to make more money, is to remove all of the wisdom teeth, when often times patients do not need all of...*
them removed. It's just like, we're in there, so let's do them all. [User #1106]

Delivering or Forgoing Dental Care

Tweets about delaying or forgoing dental care due to an inability to afford treatment were categorized into the second most prominent theme. Strategies resulting from delaying or forgoing dental care mentioned within this theme included dental tourism, visits to the emergency room, visits to free clinics or dental schools, and self-treatment.

Users tweeted about not being able to afford the upfront cost of care and as a result delaying care despite being in pain. In some cases, users shared experiences of death resulting from an inability to access dental care.

@username I had a friend who didn't get treated, he died from sepsis. I don't want to scare you but it can be infected and that can spread really fast. No dentist should refuse someone because they can't pay when it comes to infection. I wish her well. [User #1126]

Users reported visiting the emergency room to address dental pain or infection when they could not afford treatment by a dentist. They often expressed dissatisfaction with their experience, which often left them with palliative care such as pain killers and antibiotics rather than treatment. In some cases, visits to the emergency room resulted in unexpected costs when medical insurances denied claims for dental-related conditions.

@username will not cover the ER visit because dental related, and I will have to pay close to $2000 hospital bill. All they did was examine me and prescribe painkiller and antibiotics. How is this right? Everything needs to change. [User #4995]

Travel to countries with lower-cost dental care was expressed as a way to access dental care when dental treatment in the United States was considered to be too expensive. Domestically, users considered going to dental schools or public dental clinics when they could not afford the cost of a private dentist.

The fact that I’m driving three hours into Mexico tomorrow to get my teeth done at an 1/8 of the price I was going to pay at a “nonprofit” dentist in America is fucking ridiculous. [User #1990]

Payment Strategies

The theme about “payment strategies” for needed dental care most frequently included tweets asking for donations from others. Certain twitter accounts were frequently referenced in donation requests including @pulte, @TeamPulte, and @JefferyStar. Twitter users often asked for funds to be sent to them via PayPal, Venmo, GoFundMe, and CashApp.

I’m raising money to have my teeth pulled. Click to Donate: [URL] via @gofundme. [User #241]

A limited number of tweets indicated having to make sacrifices, such as foregoing groceries or rent, to afford dental care. Other payment strategies included going into debt and using stimulus checks or tax refunds to pay for care.

Really hoping I get stimulus money. I have teeth that hurt and dentists are expensive. [User #1091]

Insurance

Tweets discussing experiences with private-payer dental insurance were included in this less prominent theme. Many tweets discussed dental insurance as an employee benefit, and some expressed gratitude for occupation-granted insurance facilitating access to dental care.

However, tweets about insurance for dental care expressed negative sentiments more frequently, including a fear of losing coverage or not being able to afford care despite having insurance. In certain instances, individuals expressed displeasure with the complexity of the insurance system and dissatisfaction for having to pay for a portion of dental care in addition to their monthly insurance premium. Some users also expressed displeasure with dental care being excluded from medical insurance.

@username I have fancy employee dental insurance, but still have to pay a bunch for crowns and root canals with the added bonus of being really limited in dentists who accept the insurance. Instead of paying $1,000 for a crown I pay $300. So basically paying $30/month for a discount card. [User #4934]

Policy Statements

Tweets categorized as “policy statements” were the least prominent theme. The included tweets were those that mentioned Medicaid or Medicare, health care reform, were directed at politicians, or were about veterans. The content of these tweets was often related to the desired policy changes related to dental coverage.

Time and Geographic Distribution of Tweets

Overall, 5314 tweets about dental affordability were collected (Table 1), with 277 (5.2%) in 2017, 1227 (23.1%) in 2018, 1341 (25.2%) in 2019, 1090 (20%) in 2020, and 1379 (26%) in 2021. The period of data collection was less than a year in 2017 (September 1 to December 31; 4 months) and 2021 (January 1 to September 30; 9 months).

The volume of tweets over the study period is depicted in Figure 2. There was a monthly average of 116 included tweets. The lowest number of tweets was observed in April 2020 (58 tweets, 1.1%), and the highest number of tweets was observed in September 2021 (673 tweets, 12.7%). Overall and across each year of collection, “general sentiments” was the most prevalent thematic category (Table 1). “Delaying or foregoing care” was the second most prevalent thematic category, except in 2019 and 2021, when “payment strategies” and “policy statements,” respectively, were the more frequently observed categories. “Policy statements” was consistently the least frequently observed thematic category, except for a significant increase in 2021.

The geographic location of tweets was determined and mapped in Figure 3. Tweets originated the most from the most populous states in the United States, which are California, Texas, Florida, and New York [34].
Discussion

Principal Results and Comparisons With Prior Work
This study leveraged Twitter to examine public sentiments toward dental care affordability in the United States. Twitter users expressed dissatisfaction with the cost of dental treatment and the ability to access dental care. The overwhelmingly negative sentiments found in this study provide insight into how individuals are coping with financial barriers, including delaying or foregoing care and pursuing various payment strategies.
study findings support the conclusions of the existing dental literature on oral care affordability [2-5,7]. In the 2019 National Health Interview Survey, it was found that 19.2% of women and 15.6% of men did not access their needed dental care because of cost in the prior 12 months [35]. Rates of forgoing dental care are particularly high for uninsured adults, where 1 in 2 had not seen a dentist because of costs [36]. Moreover, our study findings highlight the fact that financial barriers to accessing dental care are prevalent even among those with dental insurance. The expressed sentiments of discontent with dental insurance payment structures echoed those of the existing reports in the literature. In a study on US health care spending, among 154 conditions examined, oral disorders requiring dental care had the highest out-of-pocket costs [37]. Another study that supports the sentiments of insured tweeters in this study reported that among individuals who were insured all year, US adults were significantly more likely than adults in other developed countries to go without care because of costs, the possibility of facing high out-of-pocket spending, or the financial burden of medical bills [36]. In an effort to come up with the money required to be paid out of pocket, users reported a variety of payment strategies that included credit cards, loans, stimulus checks, tax refunds, and donations. Crowdfunding, the web-based solicitation of public donations, has become a major financer of health care–related costs [38-41]. This trend was reflected in our study findings and reinforces sentiments that the use of crowdfunding to cover direct health care expenses may be a sign of a failing system [42]. Future research may explore crowdfunding for dental procedures.

Two relevant societal events occurred during the study period, which were the COVID-19 pandemic and the election of President Biden. COVID-19 resulted in dental office closures across the country in the spring and summer of 2020, with the greatest decline in weekly visits compared with 2019 observed in the week of April 12, 2020 [43]. This trend corresponds with the quantitative findings of this study, in which the lowest number of tweets were observed in April 2020. Further, COVID-19 had an economic impact that appeared to exacerbate financial barriers to accessing dental care for some people. One user tweeted, “…I can’t wait to get an eye exam and new glasses, not to mention long overdue dental care! Or just not because I can’t afford it … because of #COVID19”. For others, economic stimulus checks distributed during the COVID-19 pandemic helped reduce financial barriers to dental care; for example, “@politico Got my stimulus money today. Thinking of getting much-needed dental work. Several teeth fell out when I had COVID in Dec. and Jan. Thank you President Joe!” Lastly, once dental offices reopened after COVID-19–related closures, there were tweets expressing displeasure with increased costs of care due to safety measures; for example, “…Had to pay an up charge for PPE for a dental appointment over the summer. Was told insurance was not likely to cover that aspect of the cleaning…”. The election of President Biden in 2021 renewed conversations about health care reform, and the volume of tweets coded as “policy statements” increased in correspondence. In September 2021, there were 3.4 times the monthly average number of tweets. We hypothesize that this increase is a reaction to the announcement in August 2021 that President Joe Biden’s budget-reconciliation package included funding for a standard Medicare dental benefit [44-46]. Ultimately, the Medicare dental benefit was not included in the House of Representatives’ passed legislation.

Limitations
The trends in the volume of tweets over time as well as the geographic distribution of the tweets support the generalizability and transferability of the findings of this study. However, this study is not without limitations. While all demographic categories have been shown to engage with social media to varying extents in the United States, Twitter users are not necessarily representative of the US population [47]. On average, Twitter users are younger, are more likely to identify as Democrats, are more highly educated, and have higher incomes than US adults overall [48]. Further, the sample of tweets collected may have some sampling bias, as the data set was not necessarily limited to a single tweet per unique user.

Conclusions
The findings illustrate the real-time sentiment of Twitter users toward the cost of dental treatment and suggest shortcomings in funding, which may be representative of a greater systemic failure in the provision of dental care. Thus, this study provides insights for policy makers and dental professionals who strive to increase access to dental care. Limitations of dental insurance payment models, both public and private, are one such area that may be explored.

Acknowledgments
The authors acknowledge Jennifer Gibbs, MAS, DDS, PhD for her guidance on the pilot study for this project.

Conflicts of Interest
None declared.

References


©Shahen Yashpal, Ananditha Raghunath, Nihan Gencerliler, Lorel E Burns. Originally published in JMIR Formative Research (https://formative.jmir.org), 01.07.2022. 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.