Original Papers

A Conversational Artificial Intelligence Agent for a Mental Health Care App: Evaluation Study of Its Participatory Design (e30053)
Morena Danieli, Tommaso Ciulli, Seyed Mousavi, Giuseppe Riccardi ................................................................. 4

An Electronic Patient-Reported Outcome Mobile App for Data Collection in Type A Hemophilia: Design and Usability Study (e25071)
Francesco Petracca, Rosaria Tempre, Maria Cuccinello, Oriana Ciani, Elena Pompeo, Luigi Sannino, Valeria Lovato, Giancarlo Castaman, Alessandra Ghirardini, Rosanna Tarricone ................................................................. 20

Use and Appreciation of a Web-Based, Computer-Tailored Diet and Physical Activity Intervention Based on the Self-determination Theory: Evaluation Study of Process and Predictors (e22390)
Juul Coumans, Anke Oenema, Catherine Bolman, Lilian Lechner ........................................................................ 32

Development of a Web-Based School Support System Within the AVATAR Project for Psychosocial Well-being in Adolescents: Pilot Feasibility Study (e24840)
Francesca Mastorci, Paolo Piaggi, Gabriele Trivellini, Cristina Doveri, Anselmo Casu, Luca Bastiani, Irene Marinaro, Cristina Vassalle, Alessandro Pingitore ........................................................................................................ 53

Development of the Shift Smartphone App to Support the Emotional Well-Being of Junior Physicians: Design of a Prototype and Results of Usability and Acceptability Testing (e26370)
Isabelle Counson, Alexandra Bartholomew, Joanna Crawford, Katherine Petrie, Geetanjali Basakod, Victoria Moynihan, Josie Pires, Rachel Cohen, Nicholas Glosier, Samuel Harvey, Samineh Sanatkar ......................................................................................... 63

The Use of Food Images and Crowdsourcing to Capture Real-time Eating Behaviors: Acceptability and Usability Study (e27512)
Katharine Harrington, Shannon Zenk, Linda Van Horn, Lauren Giurini, Nithya Mahakala, Kiari Kershaw ................................................................................................................................. 77

Users' Perceptions Toward mHealth Technologies for Health and Well-being Monitoring in Pregnancy Care: Qualitative Interview Study (e28628)
Jane Li, David Silvera-Tawil, Marlien Varnfield, M Hussain, Vanitha Math ........................................................................ 89

Initial Feasibility of the “Families Moving Forward Connect” Mobile Health Intervention for Caregivers of Children With Fetal Alcohol Spectrum Disorders: Mixed Method Evaluation Within a Systematic User-Centered Design Approach (e29887)
Christie Petrenko, Carson Kautz-Turnbull, Alicia Roth, Jennifer Parr, Cristiano Tapparello, Utku Demir, Heather Olson ................. 103

Conversational Agent for Healthy Lifestyle Behavior Change: Web-Based Feasibility Study (e27956)
Dhakshenya Dhinagaran, Thirunavukkarasu Sathish, AlJia Soong, Yin-Leng Theng, James Best, Lorainne Tudor Car ................................................. 127
Patient-Reported Outcome Measure for Real-time Symptom Assessment in Women With Endometriosis: Focus Group Study (e28782)
Esther van Barneveld, Arianne Lim, Nehalennia van Hanegem, Lisa Vork, Alexandra Herrewegh, Mikal van Poll, Jessica Manders, Frits van Osch, Wilbert Spaans, Gommert van Koeveringe, Desiree Vrijens, Joanna Kruijmel, Marlies Bongers, Carsten Leue. 143

Assessing the Views of Professionals, Patients, and Care Partners Concerning the Use of Computer Tools in Memory Clinics: International Survey Study (e31053)
Aniek van Gils, Leonie Visser, Heleen Hendriksen, Jean Georges, Majon Muller, Femke Bouwman, Wiesje van der Flier, Hanneke Rhodius-Meester. 154

A Multimodal Messaging App (MAAN) for Adults With Autism Spectrum Disorder: Mixed Methods Evaluation Study (e33123)
Mohamad Hijab, Dena Al-Thani, Bilikis Banire. 168

Assessing Health Data Security Risks in Global Health Partnerships: Development of a Conceptual Framework (e25833)
Juan Espinoza, Abu Sikder, James Dickhoner, Thomas Lee. 182

Test-Retest Reliability of Home-Based Fitness Assessments Using a Mobile App (R Plus Health) in Healthy Adults: Prospective Quantitative Study (e28040)
I-I Lin, You-Lin Chen, Li-Ling Chuang. 197

Adoption and Appropriateness of mHealth for Weight Management in the Real World: A Qualitative Investigation of Patient Perspectives (e29916)
Jessica Breland, Khizran Agha, Rakshitha Mohankumar. 214

Predicting Risk of Stroke From Lab Tests Using Machine Learning Algorithms: Development and Evaluation of Prediction Models (e23440)
Eman Alanazi, Aalaa Abdou, Jake Luo. 223

Using Artificial Neural Network Condensation to Facilitate Adaptation of Machine Learning in Medical Settings by Reducing Computational Burden: Model Design and Evaluation Study (e20767)
Dianbo Liu, Ming Zheng, Nestor Sepulveda. 233

Transitioning From In-Person to Remote Clinical Research on Depression and Traumatic Brain Injury During the COVID-19 Pandemic: Study Modifications and Preliminary Feasibility From a Randomized Controlled Pilot Study (e28734)
Lauren Fisher, Sylvie Tuchman, Andrew Curreri, Maggie Markgraf, Maren Nyer, Paolo Cassano, Grant Iverson, Maurizio Fava, Ross Zafonte, Paola Pedrelli. 242

Depression, Anxiety, and Daily Activity Among Adolescents Before and During the COVID-19 Pandemic: Cross-sectional Survey Study (e30702)
Anna Jolliff, Qianqian Zhao, Jens Eickhoff, Megan Moreno. 253

Pandemic Information Dissemination and Its Associations With the Symptoms of Mental Distress During the COVID-19 Pandemic: Cross-sectional Study (e28239)
Ole Amundsen, Asle Hoffart, Sverre Johnson, Omid Ebrahimi. 263

Re-examining COVID-19 Self-Reported Symptom Tracking Programs in the United States: Updated Framework Synthesis (e31271)
Miranda Janvrin, Jessica Korona-Bailey, Tracey Koehlmoos. 276
Impact of the COVID-19 Pandemic on a Physician Group’s WhatsApp Chat: Qualitative Content Analysis
(e31791)
Sawsan Abdel-Razig, Pascale Anglade, Halah Ibrahim.
Original Paper

A Conversational Artificial Intelligence Agent for a Mental Health Care App: Evaluation Study of Its Participatory Design

Morena Danieli1*, PsyD; Tommaso Ciulli2, PhD; Seyed Mahed Mousavi1*, MSc; Giuseppe Riccardi1*, PhD

1Speech and Interactive Signal Lab, Department of Engineering and Computer Science, Università degli Studi di Trento, Trento, Italy
2IDEGO srl, Rome, Italy
*these authors contributed equally

Corresponding Author:
Morena Danieli, PsyD
Speech and Interactive Signal Lab
Department of Engineering and Computer Science
Università degli Studi di Trento
Via Sommarive 5, Povo di Trento
Trento, 38123
Italy
Phone: 39 0461282087
Email: morena.danieli@unitn.it

Abstract

Background: Mobile apps for mental health are available on the market. Although they seem to be promising for improving the accessibility of mental health care, little is known about their acceptance, design methodology, evaluation, and integration into psychotherapy protocols. This makes it difficult for health care professionals to judge whether these apps may help them and their patients.

Objective: Our aim is to describe and evaluate a protocol for the participatory design of mobile apps for mental health. In this study, participants and psychotherapists are engaged in the early phases of the design and development of the app empowered by conversational artificial intelligence (AI). The app supports interventions for stress management training based on cognitive behavioral theory.

Methods: A total of 21 participants aged 33-61 years with mild to moderate levels of stress, anxiety, and depression (assessed by administering the Italian versions of the Symptom Checklist-90-Revised, Occupational Stress Indicator, and Perceived Stress Scale) were assigned randomly to 2 groups, A and B. Both groups received stress management training sessions along with cognitive behavioral treatment, but only participants assigned to group A received support through a mobile personal health care agent, designed for mental care and empowered by AI techniques. Psychopathological outcomes were assessed at baseline (T1), after 8 weeks of treatment (T2), and 3 months after treatment (T3). Focus groups with psychotherapists who administered the therapy were held after treatment to collect their impressions and suggestions.

Results: Although the intergroup statistical analysis showed that group B participants could rely on better coping strategies, group A participants reported significant improvements in obsessivity and compulsivity and positive distress symptom assessment. The psychotherapists’ acceptance of the protocol was good. In particular, they were in favor of integrating an AI-based mental health app into their practice because they could appreciate the increased engagement of patients in pursuing their therapy goals.

Conclusions: The integration into practice of an AI-based mobile app for mental health was shown to be acceptable to both mental health professionals and users. Although it was not possible in this experiment to show that the integration of AI-based conversational technologies into traditional remote psychotherapy significantly decreased the participants’ levels of stress and anxiety, the experimental results showed significant trends of reduction of symptoms in group A and their persistence over time. The mental health professionals involved in the experiment reported interest in, and acceptance of, the proposed technology as a promising tool to be included in a blended model of psychotherapy.

(JMIR Form Res 2021;5(12):e30053) doi:10.2196/30053

KEYWORDS
mental health care; conversational AI; mHealth; personal health care agents; participatory design; psychotherapy

https://formative.jmir.org/2021/12/e30053
Introduction

Background
During the past 10 years, a multitude of mental health apps have been made available in the market [1,2]. Their functionalities range from (1) delivering questionnaires for mood self-monitoring [3,4] and (2) providing recommendations for emotion regulation [5] to (3) engaging users in rule-based interactions [6], sometimes with the support of web-based scripted dialogs [7]. As the requirements for mental health services is widespread [8] and with the current COVID-19 pandemic creating a spike in demand (as stated by the World Health Organization surveys on October 10, 2020 [9,10]), there is a greater awareness of these apps among mental health professionals [11,12]. However, there is little consensus on the usability and effectiveness of such systems [13]. Some independent research studies observed that often there is poor engagement from patients in continuing to use the apps after a few attempts [1]; others report concerns from the point of view of security, privacy, and ethical implications [14,15].

An increasing number of review papers have studied the use of chatbots in mental health. Chatbots are an evolution of internet-mediated psychological interventions. Although the latter were developed for supporting psychological care by prescriptive models, chatbots aim to engage users in short conversations about their mental distress. In the mental health domain, chatbots are often based on scripted or Eliza-style dialogues. Bendig et al [17] have analyzed the results from 10 pilot studies published between 2009 and 2018. The goal of these pilots was to assess user acceptance and effectiveness of the therapeutic recommendations, but many of them mostly included nonclinical samples. The meta-analysis by Bendig et al supports the view that state-of-the-art mental health chatbots are still experimental and that little evidence for transferring results to real psychotherapy contexts is available. In addition, Lim and Penn [18], who studied the potential of the application of digital technology in schizophrenia therapy, have stressed the need for reliable data, and the recent review by Gaffney et al [19] has highlighted the need for relying on unbiased data. However, Gaffney et al [19] have also stressed the importance of focusing current research in this field on the identification of the key mechanisms of action of the conversational agent interventions. This is very important, and in our view this aspect may be improved by meeting 2 requirements; that is, on the one hand by basing the interaction model of conversational agents on principled theoretical explanations of psychological change and on the other hand by involving mental health professionals in the design studies of blended interventions. This paper takes both recommendations into careful consideration.

Objective
It should be noted that in the crowded landscape of mental health apps, there is a lack of principled protocols for developing personal agent-driven mental health interventions. Moreover, the involvement of mental health professionals in the design of the apps is almost missing, both in the phase of setting the requirements and in the evaluation of outcomes.

In this paper, we describe the protocol we are applying to develop Therapy Empowerment Opportunity or TEO, a mobile personal health care agent (m-PHA) for mental health whose goal is to support patients dealing with the perception of augmented levels of stress and anxiety related to problems in their workplace. In particular, the goal of our research is to test a protocol for investigating the opportunity offered by the integration of artificial intelligence (AI)–enabled conversational technology into a protocolized model of psychological treatment of work-related stress with the aim of increasing personal coping resources. Although different psychological approaches to the treatment of stress and anxiety offer important insights into the roots of burnout and work-related stress, for example, individual psychology [20] and different declinations of psychodynamic theory [21,22], we chose to integrate the m-PHA support into a protocol for the prevention and treatment of work-related stress based on cognitive behavioral theory (stress management training [SMT] and cognitive behavioral therapy [CBT]).

CBT is based on the cognitive theory concept that psychological distress is maintained by internal (cognitive) factors and activated by external factors. Emotional distress and maladaptive behavioral reactions are caused by maladaptive cognitions [23,24]. Changing cognitions and thoughts can help to reduce symptoms [25]. The effectiveness of these treatments has been proved in several studies: 4 meta-analyses showed how CBT performed better than the other interventions in the treatment of occupational stress [25].

SMT programs are widely used for therapeutic purposes, with proven effectiveness. These programs combine specific techniques such as relaxation with CBT. This approach considers stress to be the imbalance between strong demands (external or internal) and few individual coping resources. The goal of SMT interventions is to reduce the intensity of demands and increase coping resources [26]. The delivery of SMT interventions within the framework of cognitive behavioral principles has been shown to be effective for managing psychological distress related to work [27].

The approach is novel because it aims to (1) design the conversational features of the m-PHA to allow a natural and personal conversation and (2) allow the therapist to monitor patients’ progress and difficulties during the time between a session and the one that follows. For this purpose, the m-PHA engages the patients in short conversations that are not scripted but are based on the recognition of their emotional state and on the understanding of the personal content written during the period of the intervention. For example, if the user reports issues in their relationship with colleagues—“Today was a bad day because my boss asked me to complete my assignment before the conveyed deadline”—the m-PHA asks contextually appropriate questions such as “You wrote that you had a bad day with your boss due to his request to finish a task in advance of the agreed time. What emotions did you have, what mental images and thoughts?” Figure 1 represents the information flow in the system architecture.
A group of CBT therapists was involved in the process of designing this protocol as they provided information for identifying the variables that could be more suggestive of possible effectiveness of the approach. On the basis of these preliminary investigations, we set our research questions about the effectiveness (in terms of symptom reduction) of the joint use of psychotherapy and m-PHA, its possible persistence over time, and the acceptance of this integrated model by users and clinicians. The study is part of the European Union–funded Horizon 2020 research project COADAPT, whose aim is to develop methodologies to reduce work-related stress in aging workers.

**Methods**

The protocol and experimental plan were approved by the ethical committee of the University of Trento in Trento, Italy. The methodology of the intervention is described below and summarized in the CONSORT (Consolidated Standards of Reporting Trials) diagram (Figure 2).

---

**Figure 1.** The information flow in the proposed approach. The patients interact with the mobile personal health care agent (m-PHA) to share personal recollections of their life events. The therapists supervise the interaction of the m-PHA with the patients and elaborate on the patients’ personal narratives during the therapy session. ABC: antecedents, beliefs, and consequences; m-PHA: mobile personal health care agent.
Figure 2. The CONSORT (Consolidated Standards of Reporting Trials) diagram shows the flow of the intervention, the enrollment of participants, their allocation to treatment, follow-up, and analysis. CBT: cognitive behavioral therapy; m-PHA: mobile personal health care agent; SMT: stress management training; T3: assessment of psychopathological outcomes 3 months after treatment.

Recruitment

The study participants were recruited in Italy from aging workers who showed mild to high levels of distress or mild to moderate levels of anxiety and depression. The modalities for being enrolled in the study were described at psychoeducational seminars about work-related stress. A total of 160 workers participated in the seminars that were held at their workplace, of whom 64 (40%) showed interest in participating in the phases to follow of the protocol. Of these 64 workers, 29 (45%) decided to sign the informed consent forms and to undergo assessment of their levels of stress, anxiety, depression, and degree of well-being at their workplace. To select the participants, we administered the questionnaires described in the next paragraph. In addition, the participants tested negative for signs of mild cognitive impairment on the basis of the Montreal Cognitive Assessment (MoCA). The exclusion criteria included the presence of severe depression (Symptom Checklist-90-Revised [SCL-90-R] score >64), underlying psychiatric conditions, and neuropsychological mild impairment (MoCA score <26).

Description of the Questionnaires for Initial and Final Assessment

The tests used for the initial assessment (T1) were the Italian versions of the SCL-90-R [28,29], the Perceived Stress Scale (PSS) [30,31], and the Occupational Stress Indicator (OSI) [32,33]. The SCL-90-R is a 90-item self-administered questionnaire that assesses a broad spectrum of psychological problems and psychopathological symptoms, measuring both internalizing symptoms (depression, somatization, and anxiety) and externalizing symptoms (aggression, hostility, and impulsivity). The questionnaire assesses 9 primary symptom dimensions: somatization, obsessiveness-compulsiveness, interpersonal hypersensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism. There are 3 global indexes: Global Severity Index (GSI), Positive Symptom Total (PST), and Positive Symptom Distress Index (PSDI). The PSS is a widely adopted questionnaire for the measurement of psychological stress. It is a self-reported questionnaire that was designed to assess “the degree to which individuals appraise situations in their lives as stressful” [30]. The OSI is a test for the wide-ranging detection of psychosocial stress in organizations. The different sections that make up the

https://formative.jmir.org/2021/12/e30053
test detect the causes of perceived stress, their consequences, and individual coping resources. A further element detected by the instrument is the evaluation of some personal characteristics that, more than other characteristics, can promote stress. The Italian version of the MoCA was administered for assessing the absence of mild cognitive impairment [34].

Protocol

In all, 8 psychotherapy sessions with CBT therapists were held through videoconference on a weekly basis. During the first session, the patients were invited to use the m-PHA to complete the assignments they received during the sessions, which included the writing of ABC (antecedents, beliefs, and consequences) notes.

The ABC technique is used in CBT to make individuals aware of their thoughts and to help them understand the link among events (antecedents), thoughts (beliefs), and emotions and behaviors (consequences). The technique increases understanding of nonfunctional behaviors and irrational or dysfunctional beliefs. The ABC technique was initially introduced by Ellis [24] and subsequently taken up by Beck [23]. The basic theory is that it is not events (A) that directly generate certain emotions but how these events are cognitively processed and evaluated and how irrational or dysfunctional beliefs (B) influence this processing [35-37].

In this protocol, the m-PHA conversed with the users to give names to the emotions they felt, to recognize their physical manifestations, and to localize them in some part of their bodies. In addition, it could provide suggestions for doing relaxation exercises. At the end of the psychotherapy treatment (T2), the participants received the same questionnaires submitted at T1, with the exclusion of the neuropsychological assessment. After 3 months, the study participants were contacted again for the third assessment (T3). At the end of the intervention, the psychotherapists involved in the experiment were engaged in a focus group to collect their opinions about the feasibility of integrating the m-PHA into the SMT-CBT protocol they apply with their patients.

Participants

Sample characteristics are described in Table 1. A total of 29 potential participants were examined, and 21 (72%) were recruited and distributed into 2 experimental groups: group A received SMT-CBT treatment and the opportunity to use the m-PHA, whereas group B received only the SMT-CBT treatment. Of the 21 participants, 11 (52%) were assigned to group A and 10 (48%) to group B. On average, group A participants were aged 46.9 (SD 5.89) years and had 22.18 (SD 8.06) years of work experience, whereas group B participants were aged 48.7 (SD 10.21) years and had 25.30 (SD 11.59) years of work experience. Assigning participants to a control group was not planned in this experiment because the total number of participants we targeted was small and the goal of this study was to assess acceptability of the blended model of psychotherapy and the possibility of psychotherapists including an AI-enabled app in their work with patients. On the basis of the results of this study, we have planned and designed further experiments (currently running) in which a subset of participants has been assigned to a control group.

Table 1. Sample characteristics (N=21).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>47.76 (8.07)</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>4 (19)</td>
</tr>
<tr>
<td>Female</td>
<td>17 (81)</td>
</tr>
<tr>
<td><strong>Groups, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Group A</td>
<td>11 (52)</td>
</tr>
<tr>
<td>Group B</td>
<td>10 (48)</td>
</tr>
<tr>
<td><strong>Formal education, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>7 (33)</td>
</tr>
<tr>
<td>Degree</td>
<td>10 (48)</td>
</tr>
<tr>
<td>Master’s degree or PhD*</td>
<td>4 (19)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>3 (14)</td>
</tr>
<tr>
<td>Cohabiting</td>
<td>4 (19)</td>
</tr>
<tr>
<td>Married</td>
<td>13 (62)</td>
</tr>
<tr>
<td>Separated</td>
<td>1 (5)</td>
</tr>
</tbody>
</table>

*PhD: Doctor of Philosophy.
Statistical Analysis

Statistical analysis was performed using nonparametric statistics for ordinal data. In addition, by following the suggestions made by an anonymous reviewer and by Sullivan and Artino [38], a parametric independent 2-tailed t test analysis of data was performed.

The nonparametric statistical analysis applied the Mann–Whitney test to assess the differences between group A and group B for the results reported in the SCL-90-R, OSI, and PSS tests. Nonparametric within-group differences were assessed by applying the Friedman test. Wilcoxon tests were used to follow up the within-group findings.

Results

Parametric Data Analysis

Overview

Parametric data analysis (independent t test) was performed on the collected data by comparing the differences between groups A and B with respect to the results obtained in the SCL-90-R, PSS (Table 2), and OSI (Table 3) questionnaires at T1, T2, and T3. For the OSI test, we only considered the scales regarding coping strategies such as home-work relationship, social support, logic, task oriented, involvement, and time.
Table 2. Parametric analysis of differences between group A (n=11) and group B (n=10) at baseline (T1), after 8 weeks of treatment (T2), and 3 months after treatment (T3): Perceived Stress Scale (PSS) and Symptom Checklist-90-Revised tests.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Group A, mean (SD)</th>
<th>Group B, mean (SD)</th>
<th>t test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PSS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>22.09 (2.21)</td>
<td>20.40 (6.83)</td>
<td>0.75 (10.7)</td>
<td>.47</td>
</tr>
<tr>
<td>T2</td>
<td>16.55 (5.45)</td>
<td>14.80 (5.45)</td>
<td>0.73 (19)</td>
<td>.47</td>
</tr>
<tr>
<td>T3</td>
<td>18 (7.32)</td>
<td>10.29 (6.63)</td>
<td>2.22 (15)</td>
<td>.04</td>
</tr>
<tr>
<td><strong>GSI</strong>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>60.36 (6.82)</td>
<td>55.70 (7.92)</td>
<td>1.45 (19)</td>
<td>.16</td>
</tr>
<tr>
<td>T2</td>
<td>55.82 (6.82)</td>
<td>49.00 (8.60)</td>
<td>2.02 (19)</td>
<td>.06</td>
</tr>
<tr>
<td>T3</td>
<td>53.89 (8.71)</td>
<td>47.00 (6.13)</td>
<td>1.67 (13)</td>
<td>.12</td>
</tr>
<tr>
<td><strong>PST</strong>b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>62.82 (6.81)</td>
<td>58.20 (9.66)</td>
<td>1.28 (19)</td>
<td>.22</td>
</tr>
<tr>
<td>T2</td>
<td>60.73 (10.38)</td>
<td>51.00 (12.13)</td>
<td>1.98 (19)</td>
<td>.06</td>
</tr>
<tr>
<td>T3</td>
<td>57.00 (8.02)</td>
<td>48.83 (7.76)</td>
<td>1.96 (13)</td>
<td>.07</td>
</tr>
<tr>
<td><strong>PSDI</strong>c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>53.82 (5.64)</td>
<td>51.00 (8.01)</td>
<td>0.94 (19)</td>
<td>.36</td>
</tr>
<tr>
<td>T2</td>
<td>48.91 (4.57)</td>
<td>47.80 (5.31)</td>
<td>0.51 (19)</td>
<td>.61</td>
</tr>
<tr>
<td>T3</td>
<td>49.00 (6.04)</td>
<td>45.17 (4.21)</td>
<td>1.34 (13)</td>
<td>.20</td>
</tr>
<tr>
<td><strong>Somatization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>52.64 (7.45)</td>
<td>50.40 (8.51)</td>
<td>0.64 (19)</td>
<td>.53</td>
</tr>
<tr>
<td>T2</td>
<td>49.64 (6.45)</td>
<td>47.50 (9.19)</td>
<td>0.62 (19)</td>
<td>.54</td>
</tr>
<tr>
<td>T3</td>
<td>48.78 (7.17)</td>
<td>46.17 (9.56)</td>
<td>0.61 (13)</td>
<td>.56</td>
</tr>
<tr>
<td><strong>Obsessiveness-compulsiveness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>62.55 (8.95)</td>
<td>51.30 (7.63)</td>
<td>3.08 (19)</td>
<td>.006</td>
</tr>
<tr>
<td>T2</td>
<td>56.55 (8.78)</td>
<td>48.20 (9.78)</td>
<td>2.06 (19)</td>
<td>.05</td>
</tr>
<tr>
<td>T3</td>
<td>55.11 (8.04)</td>
<td>48.17 (7.52)</td>
<td>1.68 (13)</td>
<td>.12</td>
</tr>
<tr>
<td><strong>Depression</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>63.27 (8.86)</td>
<td>57.80 (8.59)</td>
<td>1.43 (19)</td>
<td>.17</td>
</tr>
<tr>
<td>T2</td>
<td>57.36 (9.29)</td>
<td>51.70 (8.41)</td>
<td>1.46 (19)</td>
<td>.16</td>
</tr>
<tr>
<td>T3</td>
<td>57.44 (11.90)</td>
<td>47.33 (6.83)</td>
<td>1.87 (13)</td>
<td>.08</td>
</tr>
<tr>
<td><strong>Anxiety</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>57.55 (10.21)</td>
<td>56.50 (9.17)</td>
<td>0.25 (19)</td>
<td>.81</td>
</tr>
<tr>
<td>T2</td>
<td>59.27 (9.09)</td>
<td>49.60 (6.15)</td>
<td>2.82 (19)</td>
<td>.01</td>
</tr>
<tr>
<td>T3</td>
<td>53.00 (6.67)</td>
<td>48.33 (3.50)</td>
<td>1.56 (13)</td>
<td>.14</td>
</tr>
</tbody>
</table>

*aGSI: Global Severity Index.

bPST: Positive Symptom Total.

PSDI: Positive Symptom Distress Index.
Table 3. Parametric analysis of differences between group A (n=11) and group B (n=10) at baseline (T1), after 8 weeks of treatment (T2), and 3 months after treatment (T3): Occupational Stress Indicator test.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Group A, mean (SD)</th>
<th>Group B, mean (SD)</th>
<th>t test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Social support</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>5.91 (1.51)</td>
<td>5.90 (0.99)</td>
<td>0.02 (19)</td>
<td>.99</td>
</tr>
<tr>
<td>T2</td>
<td>6.36 (2.16)</td>
<td>5.50 (1.35)</td>
<td>1.08 (19)</td>
<td>.29</td>
</tr>
<tr>
<td>T3</td>
<td>5.90 (2.18)</td>
<td>7.33 (1.03)</td>
<td>−1.77 (13.6)</td>
<td>.10</td>
</tr>
<tr>
<td><strong>Task oriented</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>5.45 (1.51)</td>
<td>4.90 (1.37)</td>
<td>0.88 (19)</td>
<td>.39</td>
</tr>
<tr>
<td>T2</td>
<td>5.82 (1.33)</td>
<td>6.10 (1.97)</td>
<td>−0.39 (19)</td>
<td>.70</td>
</tr>
<tr>
<td>T3</td>
<td>5.30 (1.25)</td>
<td>7.50 (1.05)</td>
<td>−3.60 (14)</td>
<td>.003</td>
</tr>
<tr>
<td><strong>Logic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>5.36 (1.96)</td>
<td>4.80 (1.99)</td>
<td>0.65 (19)</td>
<td>.52</td>
</tr>
<tr>
<td>T2</td>
<td>5.55 (0.93)</td>
<td>5.90 (1.52)</td>
<td>−0.65 (19)</td>
<td>.52</td>
</tr>
<tr>
<td>T3</td>
<td>5.50 (1.58)</td>
<td>5.67 (1.37)</td>
<td>−0.21 (14)</td>
<td>.83</td>
</tr>
<tr>
<td><strong>Home-work relationship</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>5.45 (2.16)</td>
<td>6.10 (1.79)</td>
<td>−0.74 (19)</td>
<td>.47</td>
</tr>
<tr>
<td>T2</td>
<td>5.64 (1.96)</td>
<td>6.20 (1.32)</td>
<td>−0.76 (19)</td>
<td>.45</td>
</tr>
<tr>
<td>T3</td>
<td>6.10 (1.45)</td>
<td>6.83 (1.17)</td>
<td>−1.05 (14)</td>
<td>.31</td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>5.00 (1.34)</td>
<td>5.20 (1.99)</td>
<td>−0.27 (19)</td>
<td>.79</td>
</tr>
<tr>
<td>T2</td>
<td>5.09 (1.87)</td>
<td>5.80 (1.32)</td>
<td>−0.99 (19)</td>
<td>.33</td>
</tr>
<tr>
<td>T3</td>
<td>5.10 (2.18)</td>
<td>6.50 (2.07)</td>
<td>−1.26 (14)</td>
<td>.23</td>
</tr>
<tr>
<td><strong>Involvement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>5.73 (1.95)</td>
<td>5.90 (1.10)</td>
<td>−0.25 (19)</td>
<td>.81</td>
</tr>
<tr>
<td>T2</td>
<td>5.82 (1.08)</td>
<td>6.60 (1.84)</td>
<td>0.34 (19)</td>
<td>.74</td>
</tr>
<tr>
<td>T3</td>
<td>6.00 (1.25)</td>
<td>7.67 (1.37)</td>
<td>−2.5 (14)</td>
<td>.03</td>
</tr>
</tbody>
</table>

**SCL-90-R and PSS results**

At T1, the SCL-90-R obsessivity and compulsivity levels in group A (mean 62.5, SE 2.70) were significantly different from those in group B (mean 51.3, SE 2.41; \( t_{19} = 3.08; P = .006; r = 0.56 \)).

At T2, the SCL-90-R anxiety levels in group A (mean 59.3, SE 2.74) were significantly different from those in group B (mean 49.6, SE 1.94; \( t_{19} = 2.82; P = .01; r = 0.54 \)).

At T3, PSS levels in group A (mean 18, SE 2.31) were significantly different from those in group B (mean 10.3, SE 2.50; \( t_{15} = 2.22; P = .04; r = 0.49 \)).

**OSI Coping Strategies Results**

For the OSI coping strategies, only the task-oriented and involvement scales at T3 were significantly different. The task-oriented levels in group A (mean 5.3, SE 1.25) were significantly different from those in group B (mean 7.5, SE 1.05; \( t_{14} = 3.60; P = .003; r = 0.69 \)). The involvement levels in group A (mean 6. SE 0.39) were significantly different from those in group B (mean 7.67, SE 0.56; \( t_{14} = 2.50; P = .02; r = 0.56 \)).

**Nonparametric Data Analysis**

**Overview**

Nonparametric data analysis (Mann–Whitney test) was performed on the collected data by comparing the differences between the groups with respect to the results obtained in SCL-90-R, PSS (Table 4), and OSI (Table 5) questionnaires at T1, T2, and T3. For the OSI test, we only considered the scales regarding coping strategies such as home-work relationship, social support, logic, task oriented, involvement, and time.
Table 4. Nonparametric analysis of differences between group A (n=11) and group B (n=10) at baseline (T1), after 8 weeks of treatment (T2), and 3 months after treatment (T3): Perceived Stress Scale (PSS) and Symptom Checklist-90-Revised tests.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Group A Mean (SD)</th>
<th>Median</th>
<th>Mean rank</th>
<th>Group B Mean (SD)</th>
<th>Median</th>
<th>Mean rank</th>
<th>U value</th>
<th>Z value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PSS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>22.09 (2.21)</td>
<td>23.00</td>
<td>11.50</td>
<td>20.40 (6.83)</td>
<td>21.00</td>
<td>10.45</td>
<td>49.50</td>
<td>−0.39</td>
<td>.72</td>
</tr>
<tr>
<td>T2</td>
<td>16.55 (5.45)</td>
<td>17.00</td>
<td>11.73</td>
<td>14.80 (5.45)</td>
<td>15.00</td>
<td>10.20</td>
<td>47.00</td>
<td>−0.56</td>
<td>.59</td>
</tr>
<tr>
<td>T3</td>
<td>18 (7.32)</td>
<td>18.50</td>
<td>11.15</td>
<td>10.29 (6.63)</td>
<td>13.00</td>
<td>5.93</td>
<td>13.05</td>
<td>−2.10</td>
<td>.03</td>
</tr>
<tr>
<td><strong>GST</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>60.36 (6.82)</td>
<td>62.00</td>
<td>12.82</td>
<td>55.70 (7.92)</td>
<td>58.00</td>
<td>9.00</td>
<td>35.00</td>
<td>−1.41</td>
<td>.17</td>
</tr>
<tr>
<td>T2</td>
<td>55.82 (6.82)</td>
<td>58.00</td>
<td>13.59</td>
<td>49.00 (8.60)</td>
<td>49.00</td>
<td>8.15</td>
<td>26.50</td>
<td>−2.01</td>
<td>.04</td>
</tr>
<tr>
<td>T3</td>
<td>53.89 (8.71)</td>
<td>52.00</td>
<td>9.67</td>
<td>47.00 (6.13)</td>
<td>45.00</td>
<td>5.50</td>
<td>12.00</td>
<td>−1.77</td>
<td>.09</td>
</tr>
<tr>
<td><strong>PST</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>62.82 (6.81)</td>
<td>63.00</td>
<td>12.45</td>
<td>58.20 (9.66)</td>
<td>59.50</td>
<td>9.40</td>
<td>39.00</td>
<td>−1.13</td>
<td>.28</td>
</tr>
<tr>
<td>T2</td>
<td>60.73 (10.38)</td>
<td>62.00</td>
<td>13.55</td>
<td>51.00 (12.13)</td>
<td>50.50</td>
<td>8.20</td>
<td>27.00</td>
<td>−1.98</td>
<td>.05</td>
</tr>
<tr>
<td>T3</td>
<td>57.00 (8.02)</td>
<td>58.00</td>
<td>9.67</td>
<td>48.83 (7.76)</td>
<td>44.00</td>
<td>5.50</td>
<td>12.00</td>
<td>−1.77</td>
<td>.08</td>
</tr>
<tr>
<td><strong>Obsessiveness-compulsiveness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>62.55 (8.95)</td>
<td>63.00</td>
<td>13.95</td>
<td>51.30 (7.63)</td>
<td>53.00</td>
<td>7.75</td>
<td>22.50</td>
<td>−2.30</td>
<td>.02</td>
</tr>
<tr>
<td>T2</td>
<td>56.55 (8.78)</td>
<td>56.00</td>
<td>13.68</td>
<td>48.20 (9.78)</td>
<td>47.00</td>
<td>8.05</td>
<td>25.5</td>
<td>−2.08</td>
<td>.04</td>
</tr>
<tr>
<td>T3</td>
<td>55.11 (8.04)</td>
<td>54.00</td>
<td>9.50</td>
<td>48.17 (7.52)</td>
<td>46.00</td>
<td>5.75</td>
<td>13.50</td>
<td>−1.60</td>
<td>.13</td>
</tr>
<tr>
<td><strong>Depression</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>63.27 (8.86)</td>
<td>64.00</td>
<td>12.73</td>
<td>57.80 (8.59)</td>
<td>59.00</td>
<td>9.10</td>
<td>36.00</td>
<td>−1.34</td>
<td>.19</td>
</tr>
<tr>
<td>T2</td>
<td>57.36 (9.29)</td>
<td>57.00</td>
<td>12.68</td>
<td>51.70 (8.41)</td>
<td>48.50</td>
<td>9.15</td>
<td>36.50</td>
<td>−1.31</td>
<td>.20</td>
</tr>
<tr>
<td>T3</td>
<td>57.44 (11.90)</td>
<td>55.00</td>
<td>9.94</td>
<td>47.33 (6.83)</td>
<td>44.50</td>
<td>5.08</td>
<td>9.50</td>
<td>−2.07</td>
<td>.04</td>
</tr>
<tr>
<td><strong>Anxiety</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>57.55 (10.21)</td>
<td>54.00</td>
<td>10.82</td>
<td>56.50 (9.17)</td>
<td>56.00</td>
<td>11.20</td>
<td>53.00</td>
<td>−0.14</td>
<td>.90</td>
</tr>
<tr>
<td>T2</td>
<td>59.27 (9.09)</td>
<td>62.00</td>
<td>14.18</td>
<td>49.60 (6.15)</td>
<td>49.00</td>
<td>7.50</td>
<td>20.00</td>
<td>−2.48</td>
<td>.01</td>
</tr>
<tr>
<td>T3</td>
<td>53.00 (6.67)</td>
<td>53.00</td>
<td>9.39</td>
<td>48.33 (3.50)</td>
<td>47.00</td>
<td>5.92</td>
<td>14.50</td>
<td>−1.48</td>
<td>.15</td>
</tr>
</tbody>
</table>

*aGSI: Global Severity Index.

bPST: Positive Symptom Total.
Table 5. Nonparametric analysis of differences between group A (n=11) and group B (n=10) at baseline (T1), after 8 weeks of treatment (T2), and 3 months after treatment (T3): Occupational Stress Indicator test.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Group A</th>
<th>Median</th>
<th>Mean rank</th>
<th>Group B</th>
<th>Median</th>
<th>Mean rank</th>
<th>U value</th>
<th>Z value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Social support</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>5.91 (1.51)</td>
<td>6.00</td>
<td>11.05</td>
<td>5.90 (0.99)</td>
<td>6.00</td>
<td>10.95</td>
<td>54.5</td>
<td>−0.04</td>
<td>.99</td>
</tr>
<tr>
<td>T2</td>
<td>6.36 (2.16)</td>
<td>7.00</td>
<td>12.95</td>
<td>5.50 (1.35)</td>
<td>6.00</td>
<td>8.85</td>
<td>33.5</td>
<td>−1.56</td>
<td>.13</td>
</tr>
<tr>
<td>T3</td>
<td>5.90 (2.18)</td>
<td>6.5</td>
<td>7.60</td>
<td>7.33 (1.03)</td>
<td>7.00</td>
<td>10.00</td>
<td>21.00</td>
<td>−1.00</td>
<td>.32</td>
</tr>
<tr>
<td><strong>Task oriented</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>5.45 (1.51)</td>
<td>5.00</td>
<td>12.05</td>
<td>4.90 (1.37)</td>
<td>5.00</td>
<td>9.85</td>
<td>43.5</td>
<td>−0.84</td>
<td>.44</td>
</tr>
<tr>
<td>T2</td>
<td>5.82 (1.33)</td>
<td>6.00</td>
<td>10.14</td>
<td>6.10 (1.97)</td>
<td>6.00</td>
<td>11.95</td>
<td>45.50</td>
<td>−0.68</td>
<td>.53</td>
</tr>
<tr>
<td>T3</td>
<td>5.30 (1.25)</td>
<td>6.00</td>
<td>5.95</td>
<td>7.50 (1.05)</td>
<td>7.00</td>
<td>12.75</td>
<td>4.50</td>
<td>−2.85</td>
<td>.004</td>
</tr>
<tr>
<td><strong>Logic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>5.36 (1.96)</td>
<td>6.00</td>
<td>11.91</td>
<td>4.80 (1.99)</td>
<td>5.00</td>
<td>10.00</td>
<td>45.00</td>
<td>−0.73</td>
<td>.49</td>
</tr>
<tr>
<td>T2</td>
<td>5.55 (0.93)</td>
<td>5.00</td>
<td>10.09</td>
<td>5.90 (1.52)</td>
<td>6.00</td>
<td>12.00</td>
<td>45.00</td>
<td>−0.73</td>
<td>.5</td>
</tr>
<tr>
<td>T3</td>
<td>5.50 (1.58)</td>
<td>5.5</td>
<td>8.30</td>
<td>5.67 (1.37)</td>
<td>6.00</td>
<td>8.83</td>
<td>28.00</td>
<td>−0.22</td>
<td>.85</td>
</tr>
<tr>
<td><strong>Home-work relationship</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>5.45 (2.16)</td>
<td>5.00</td>
<td>10.05</td>
<td>6.10 (1.79)</td>
<td>6.00</td>
<td>12.05</td>
<td>44.5</td>
<td>−0.75</td>
<td>.46</td>
</tr>
<tr>
<td>T2</td>
<td>5.64 (1.96)</td>
<td>6.00</td>
<td>10.27</td>
<td>6.20 (1.32)</td>
<td>7.00</td>
<td>11.80</td>
<td>47.00</td>
<td>−0.58</td>
<td>.58</td>
</tr>
<tr>
<td>T3</td>
<td>6.10 (1.45)</td>
<td>6.00</td>
<td>7.60</td>
<td>6.83 (1.17)</td>
<td>7.00</td>
<td>10.00</td>
<td>21.00</td>
<td>−1.00</td>
<td>.35</td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>5.00 (1.34)</td>
<td>6.00</td>
<td>10.86</td>
<td>5.20 (1.99)</td>
<td>5.00</td>
<td>11.15</td>
<td>53.5</td>
<td>−0.11</td>
<td>.95</td>
</tr>
<tr>
<td>T2</td>
<td>5.09 (1.87)</td>
<td>6.00</td>
<td>9.50</td>
<td>5.80 (1.32)</td>
<td>6.00</td>
<td>12.65</td>
<td>38.50</td>
<td>−1.23</td>
<td>.22</td>
</tr>
<tr>
<td>T3</td>
<td>5.10 (2.18)</td>
<td>5.00</td>
<td>7.40</td>
<td>6.50 (2.07)</td>
<td>6.50</td>
<td>10.33</td>
<td>19.00</td>
<td>−1.22</td>
<td>.23</td>
</tr>
<tr>
<td><strong>Involvement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>5.73 (1.95)</td>
<td>6.00</td>
<td>10.23</td>
<td>5.90 (1.10)</td>
<td>6.00</td>
<td>11.85</td>
<td>46.5</td>
<td>−0.63</td>
<td>.56</td>
</tr>
<tr>
<td>T2</td>
<td>5.82 (1.08)</td>
<td>6.00</td>
<td>11.09</td>
<td>5.60 (1.84)</td>
<td>6.00</td>
<td>10.90</td>
<td>54.00</td>
<td>−0.07</td>
<td>.95</td>
</tr>
<tr>
<td>T3</td>
<td>6.00 (1.25)</td>
<td>6.00</td>
<td>6.55</td>
<td>7.67 (1.37)</td>
<td>7.50</td>
<td>11.75</td>
<td>10.50</td>
<td>−2.19</td>
<td>.03</td>
</tr>
</tbody>
</table>

**SCL-90-R and PSS results**

At T1, the SCL-90-R obsessivity and compulsivity levels in group A (median 63) were significantly different from those in group B (median 53; U=22.5; Z=−2.30; P=.02; r=−0.50).

At T2, other significant differences between group A and group B were observed. With respect to the GSI levels, group A (median 58) differed from group B (median 49; U=26.5; Z=−2.01; P=.04; r=−0.44). The subscale measuring the PST of group A (median 62) differed from that of group B (median 51; U=27; Z=−1.98; P=.05; r=−0.43). With respect to anxiety, group A (median 62) differed from group B (median 49; U=20; Z=−2.48; P=.01; r=−0.54), and with respect to obsessivity and compulsivity, group A (median 56) differed from group B (median 47; U=25.5; Z=−2.08; P=.03; r=−0.45).

At T3, the depression level in group A (median 55) differed from that in group B (median 44.5; U=28.5; Z=−2.07; P=.04; r=−0.45). For the PSS test, only at T3 did the levels reported by group A (median 18.5) differ significantly from those reported by group B (median 13; U=13.5; Z=−2.10; P=.03; r=−0.46).

In summary, data analysis at T1 did not show any significant difference for the PSS and SCL-90-R tests between groups A and B, with the exception of the subscale obsessiveness-compulsiveness of the SCL-90-R test (lower levels are better; see Table 2). At T2 and T3 for the SCL-90-R test, data analysis showed some differences between the 2 groups. Participants assigned to group A seemed to report lower improvements (lower levels are better) than those assigned to group B at T2 for the GSI, PST, obsessiveness-compulsiveness, and anxiety scales and at T3 for the depression scale. For the PSS test, group B showed significant improvements (lower levels are better) than group A at T3 (Table 4).

**OSI Coping Strategies Results**

For the subscales of the OSI test, the task-oriented level in group A (median 6) was significantly different from that in group B (median 7; U=4.5; Z=−2.85; P=.004; r=−0.62), and the involvement level in group A (median 6) was also significantly different from that in group B (median 7.5; U=10.5; Z=−2.19;
Participants significantly changed over the 3 measures at T1, T2, and T3.

**Group A Within-Group Analysis**

**Parametric Data Analysis**

A parametric data analysis (1-way repeated measures analysis of variance) was performed for comparing the different results reported in the participants in group A at T1, T2, and T3.

The level of PSS ($F_{2,16}=3.25; P=0.06$) and some SCL-90-R subscales (GSI, $F_{2,16}=2.80; P=0.09$; PST, $F_{2,16}=1.58; P=0.24$; somatization, $F_{2,1}=1.44; P=0.27$; interpersonal hypersensitivity, $F_{2,1}=0.95; P=0.41$; depression, $F_{2,1}=2.34; P=0.13$; anxiety, $F_{2,1}=1.05; P=0.37$; hostility, $F_{2,1}=0.43; P=0.65$; phobic anxiety, $F_{2,1}=0.13; P=0.35$; paranoid ideation, $F_{2,1}=1.26; P=0.31$; and psychotism, $F_{2,1}=1.47; P=0.26$) did not significantly change over the 3 measures at T1, T2, and T3.

For the PSDI and obsessiveness-compulsiveness subscales of the SCL-90-R test, the results show significant change over time (PSDI, $F_{2,16}=6.47; P=0.03$, with moderate effect size $\eta^2=0.35$ and obsessiveness-compulsiveness, $F_{2,16}=6.58; P=0.008$, with large effect size $\eta^2=0.49$).

The level of the examined OSI subscales of participants did not significantly change over the 3 measures at T1, T2, and T3 (social support, $F_{2,1}=0.44; P=0.65$; task oriented, $F_{2,1}=0.49; P=0.62$; logic, $F_{2,1}=0.09; P=0.92$; home-work relationship, $F_{2,1}=1.03; P=0.37$; time, $F_{2,1}=0.04; P=0.96$; and involvement, $F_{2,1}=0.22; P=0.80$).

**Nonparametric Data Analysis**

A nonparametric data analysis was performed using the Friedman test (Pereira et al [39]) for comparing the different results reported in the participants in group A at T1, T2, and T3. The level of PSS (PSS, $\chi^2=5.3; P=0.07$) and some SCL-90-R subscales (PST, $\chi^2=4.2; P=0.15$; PSDI, $\chi^2=4.2; P=0.14$; somatization, $\chi^2=3.5; P=0.20$; interpersonal hypersensitivity, $\chi^2=0.8; P=0.71$; depression, $\chi^2=5.4; P=0.08$; anxiety, $\chi^2=1.8; P=0.45$; hostility, $\chi^2=0.8; P=0.71$; phobic anxiety, $\chi^2=2.3; P=0.33$; paranoid ideation, $\chi^2=1.7; P=0.47$; and psychotism, $\chi^2=3.0; P=0.25$) did not significantly change over the 3 measures at T1, T2, and T3.

The GSI and obsessiveness-compulsiveness subscales of participants significantly changed over the 3 measures at T1, T2, and T3 ($\chi^2=6.4; P=0.04$; $\chi^2=6.4; P=0.04$; $\chi^2=6.5; P=0.03$, respectively). Wilcoxon tests were used to follow up this finding. A Bonferroni correction was applied; therefore, all effects have been reported at a 0.0167 level of significance. It seemed that the GSI did not significantly change from T1 to T2 ($r=11.50; Z=–1.92; P=0.06$), from T1 to T3 (T=8; Z=–1.72; $P=0.09$), or from T2 to T3 (r=13; Z=–0.169; P=0.93). The obsessiveness-compulsiveness levels did not significantly change from T1 to T2 (r=11; Z=–1.96; P=0.05) or from T2 to T3 (r=20; Z=–0.30; $P=0.79$), but there was a significant change from T1 (median 63) to T3 (median 54; T=0; $Z=–2.37; P=0.02$).

In summary, in group A, only the obsessiveness-compulsiveness levels showed a significant decrease at T1 in comparison with T3. The levels of the examined OSI subscales of participants did not significantly change over the 3 measures at T1, T2, and T3 (social support, $\chi^2=0.6; P=0.77$; task oriented, $\chi^2=1.3; P=0.55$; logic, $\chi^2=0.7; P=0.76$; home-work relationship, $\chi^2=0.9; P=0.66$; time, $\chi^2=0.8; P=0.71$; and involvement, $\chi^2=0.3; P=0.90$).

**Group B Within-Group Analysis**

**Parametric Data Analysis**

A parametric data analysis (1-way repeated measures analysis of variance) was performed for comparing the different results reported in the participants in group B at T1, T2, and T3.

The level of PSS ($F_{2,13}=3.56; P=0.06$) and some SCL-90-R subscales (PSDI, $F_{2,10}=2.54; P=0.17$; obsessiveness-compulsiveness, $F_{1,5.48}=2.86; P=0.15$; interpersonal hypersensitivity, $F_{2,10}=3.85; P=0.06$; hostility, $F_{2,10}=3.71; P=0.06$; phobic anxiety, $F_{2,10}=0.147; P=0.86$; paranoid ideation, $F_{2,10}=3.20; P=0.08$; and psychotism, $F_{2,10}=2.77; P=0.11$) did not significantly change over the 3 measures, T1, T2, and T3.

For the GSI scale, the Mauchly test indicated that the assumption of sphericity had been violated, $\chi^2=7.1; P=0.03$; therefore, multivariate tests have been reported ($\eta=0.55$). The results showed significant change over time, $V=0.85, F_{2,4}=11.78; P=0.02$, with large effect size, $\eta^2=0.63$.

For the PST, depression, somatization, and anxiety subscales of the SCL-90-R test, the results showed significant change over time (PST, $F_{2,10}=8.87; P=0.006$; $\eta^2=0.64$; depression, $F_{2,10}=5.84; P=0.02$; $\eta^2=0.54$; somatization, $F_{2,10}=5.56; P=0.02$; $\eta^2=0.53$; and anxiety, $F_{2,10}=5.18; P=0.03$; $\eta^2=0.51$).

The levels of the examined OSI subscales of participants did not significantly change over the 3 measures at T1, T2, and T3 (social support, $F_{2,10}=3.89; P=0.06$; time, $F_{2,10}=2.10; P=0.17$; and home-work relationship, $F_{2,10}=3.57; P=0.70$).

For the task-oriented, logic, and involvement subscales of the OSI, the results showed significant change over time (task oriented, $F_{2,10}=7.80; P=0.009$; $\eta^2=0.61$; logic, $F_{2,10}=9.54; P=0.005$; $\eta^2=0.66$; and involvement, $F_{2,10}=5.56; P=0.02$; $\eta^2=0.59$).

**Nonparametric Data Analysis**

A nonparametric analysis of data was performed using the Friedman test (Pereira et al [39]), which allowed us to compare the different results reported in group B at T1, T2, and T3.

https://formative.jmir.org/2021/12/e30053

JMI Form Res 2021 | vol. 5 | iss. 12 | e30053 | p.14

(page number not for citation purposes)
The levels of PSS (PSS, $\chi^2=4.5$; $P=.11$) and some SCL-90-R subscales (PSDI, $\chi^2=1.7$; $P=.52$; somatization, $\chi^2=5.0$; $P=.09$; obsessional-compulsive, $\chi^2=5.3$; $P=.07$; interpersonal hypersensitivity, $\chi^2=5.7$; $P=.06$; anxiety, $\chi^2=4.7$; $P=.11$; hostility, $\chi^2=3.9$; $P=.15$; phobic anxiety, $\chi^2=1.3$; $P=.59$; paranoid ideation, $\chi^2=4.3$; $P=.14$; and psychoticism, $\chi^2=4.3$; $P=.12$) did not significantly change over the 3 measures at T1, T2, and T3. The GSI, PST, and depression subscales of participants significantly changed over the 3 measures at T1, T2, and T3 (GSI, $\chi^2=9.5$; $P=.005$; $w=0.79$; PST, $\chi^2=9.0$; $P=.008$; $w=0.75$; and depression, $\chi^2=7.9$; $P=.01$; $w=0.66$).

Wilcoxon tests were used to follow up this finding. A Bonferroni correction was applied; therefore, all effects have been reported at a 0.0167 level of significance. It seemed that GSI did not significantly change from T1 to T2 ($T=8; Z=–1.99; P=.04$), from T2 to T3 ($T=5.50; Z=–0.54; P=.60$), or from T1 to T3 ($T=0.0; Z=–2.23; P=.03$).

PST did not significantly change from T1 to T2 ($T=6; Z=–1.96; P=.05$), from T2 to T3 ($T=10.50; Z=0.0; P=.99$), and from T1 to T3 ($T=0.0; Z=–2.21; P=.03$). Depression did not significantly change from T1 to T2 ($T=10; Z=–1.79; P=.08$), from T2 to T3 ($T=9; Z=–0.31; P=.81$), and from T1 to T3 ($T=0.0; Z=–2.02; P=.06$). The levels of the examined OSI subscales of participants did not significantly change over the 3 measures at T1, T2, and T3 (home-work relationship, $\chi^2=0.5; P=.90$ and time, $\chi^2=3.4; P=.18$). The social support, task-oriented, logic, and involvement subscales significantly changed over the 3 measures at T1, T2, and T3 ($\chi^2=7.1; P=.03$; $\chi^2=8.5; P=.01$; $\chi^2=8.0; P=.01$; and $\chi^2=7.5; P=.01$, respectively).

Wilcoxon tests were used to follow up this finding. A Bonferroni correction was applied; therefore, all effects have been reported at a 0.0167 level of significance. Social support did not significantly change from T1 to T2 ($T=6.5; Z=–0.85; P=.53$), from T2 to T3 ($T=0.0; Z=–2.12; P=.06$), or from T1 to T3 ($T=0.0; Z=–1.86; P=.13$). Task oriented did not significantly change from T1 to T2 ($T=3; Z=–2.16; P=.05$), from T2 to T3 ($T=5; Z=–0.71; P=.75$), or from T1 to T3 ($T=0.0; Z=–2.32; P=.03$). Logic did not significantly change from T1 to T2 ($T=5; Z=–1.87; P=.07$), from T2 to T3 ($T=3; Z=–1.34; P=.37$), or from T1 to T3 ($T=0.0; Z=–2.12; P=.06$). Involvement did not significantly change from T1 to T2 ($T=7.5; Z=–0.65; P=.66$), from T2 to T3 ($T=2.5; Z=–1.72; P=.16$), or from T1 to T3 ($T=0.0; Z=–2.25; P=.03$).

**Qualitative Evaluation of the Intervention**

A focus group with some therapists was organized with the purpose of identifying the requirements for improving the acceptance of the m-PHA in SMT-CBT–oriented psychotherapy intervention. We chose the focus group technique because in the past this method has been found appropriate for evaluating attitudes of health care personnel, among others [19,40]. A total of 5 therapists who participated in the experiment were recruited in the group; a sixth therapist who participated in the design phase of the protocol but did not take part in the experiment played the role of facilitator. In all, 2 focus group meetings were conducted in July and September 2020. The therapists ranged in age from 29 to 39 years, the mean age being 35.05 (SD 2.40) years, and their professional experience ranged from 4 to 10 years, with a mean of 6.62 (SD 1.92) years.

The themes for the group discussion were the usefulness of including m-PHA support in the therapeutic process, their impressions about how that modification of the usual setting had an impact on the psychoeducational goals of the intervention, and the usability issues of the mobile app. Data analysis was conducted on the transcribed answers and on the notes taken during the group sessions. The data analysis was performed by following the method adopted by Berland et al [40]. The transcripts were reviewed by 2 authors (MD and TC) of this study, both with competence in conducting focus groups. From the analysis, the following relevant themes were identified.

All focus group participants reported the general impression that the m-PHA could improve patients’ engagement in their therapy goals. In the therapists’ view, the process followed for integrating this mental health mobile app into their practice was effective because the system helped their patients to complete the homework assigned by the therapists, allowing them to receive assistance while writing theirABC notes. The therapists observed that in their general practice they would usually spend more time focusing on teaching their patients how to complete their ABC notes so that they could be reviewed during the first part of the next session. In this trial, the spare time afforded to the therapists was effectively used to focus on events and related mental states that had already been shared through the app by the patients. In general, they recognized that most of the patients receiving the support of the m-PHA progressed faster in terms of the acquisition of the psychoeducational techniques of stress management.

The focus group participants carefully examined the different aspects related to the patient-therapist working alliance concerning the common goal of acquiring attitudes that may contribute to reducing the impact of stress in the patients’ everyday lives. In their view, the introduction of the m-PHA had no negative impact on the establishment of the working alliance.

As for usability issues related to the m-PHA app, the therapists expressed interest in extending the m-PHA support to their patients by including assistance in completing other types of CBT techniques, for example, disputing, in addition to the present support provided for ABC notes.

**Discussion**

**Principal Findings**

The analysis showed some significant differences between the 2 groups. The parametric analysis as well as the nonparametric analysis showed that in the examined subscales of the SCL-90-R, OSI, and SCL tests, group B seemed to show greater improvements than group A. The effect size in the parametric and nonparametric analyses was very large in scales that are significantly different.
In the SCL-90-R, for the subscales GSI, PST, anxiety, and depression, group B participants reported better changes on average than group A participants.

For the obsessivity and compulsivity scale, it is difficult to make an interpretation of what emerged because the 2 groups were different even at T1.

With respect to the PSS, group B showed better improvements than group A, especially at T3 where the effect size was very large.

As reported in the Future Research section, the conclusion of the intervention coincided with Italy entering lockdown because of COVID-19, and in the following months, different restrictions were imposed at different locations. This may have caused the increase in the level of anxiety observed in group A at T2 but not in group B, and the same circumstances applied to the level of stress at T3.

The dimensions evaluated by the OSI test, in particular the ones related to coping strategies, showed better improvements for participants assigned to group B than for group A participants. This difference was significantly different at T3 only for the task-oriented and involvement subscales, and the effect size was very large.

In addition, with regard to the mean levels of the SCL-90-R, PSS, and OSI tests, an improvement trend may be observed from T1 to T2 and from T2 to T3 in both group A and group B.

In group A, the mean of the obsessivity and compulsivity and PSDI subscales showed a significant decrease (Table 2) between assessment times, with a moderate effect for PSDI and a large effect for obsessiveness-compulsiveness. With nonparametric analysis, only the obsessiveness-compulsiveness values decreased, with a moderate effect (Table 4).

In group B, the mean of the GSI, PST, depression, somatization, and anxiety subscales showed a significant decrease (Table 2) between assessment times, with a large effect, as was the case for the task-oriented, logic, and involvement subscales, with a large effect. With nonparametric analysis, none of the SCL-90-R or OSI scales seemed to improve significantly over time, although the effect size is large. This could be an indication that sample size had an impact.

Future Research

The goal of this study is to evaluate a protocol for an intervention for the treatment of work-related stress and anxiety based on the integration of a conversational AI-empowered mobile app into traditional psychotherapy. To validate the protocol, we needed to collect data from real users to feed the machine learning algorithms of the conversational m-PHA.

More importantly, we needed to collect feedback from the psychotherapists who were involved in the participatory design effort. The limited number of participants that we could enroll did not allow the allocation of participants in more than 2 experimental groups. The research described in this paper was the initial and exploratory phase of a larger intervention protocol that is currently registered in ClinicalTrials.gov (NCT04809090). This larger protocol includes a control group, whose participants do not receive any type of treatment, as well as a fourth group, whose participants receive only the support of the m-PHA.

At the time of the data collection described in this paper, the version of the m-PHA used had limited dialog capabilities. The m-PHA was not yet able to engage participants in extended conversations: it aimed mainly to motivate users to leave personal narratives to complete the ABC homework required by the SMT-CBT protocol. The data collected in this experiment, as well as the input provided by the psychotherapists, allowed us to increase the dialog capabilities of the m-PHA.

Moreover, it is important to consider the temporal context of the data collection: the intervention phase began in December 2019 and ended in March 2020, coinciding with the first wave of the COVID-19 pandemic, and all of Italy was in lockdown for the first time. In the following months, different restrictions were imposed at different locations. During the last therapy sessions, many participants reported COVID-19–related episodes in their ABC diaries. It is likely that the participants reported anxiety levels that in some cases exceeded what they reported at the beginning of the experiment, and this was mainly because of the tragic situation that suddenly changed their daily life and, in some cases, their working conditions. In the revised protocol, the data analysis will also address the variables related to the regional variability of the COVID-19 pandemic in Italy, including the impact of regional lockdown measures.

Conclusions

The results of our study shed light on the perspectives of applying AI technologies in the field of mental health care. The goal of the work described in this paper is 2-fold. The first objective is to evaluate the intervention protocol for integrating an m-PHA into the therapeutic process. The intervention addressed work-related stress management and engaged mental health professionals in the design and test phase. This blended approach included remote sessions of traditional SMT-CBT treatment as well as the integrated support of an m-PHA. The other objective of this study is to collect natural language and behavioral data to train the machine learning algorithms of the conversational agent and to design the experimental protocol in view of the ongoing randomized controlled trial.

The results support the hypothesis that SMT-CBT treatment may be integrated into AI-based mental health agents. The therapists engaged in the participatory design model adopted in this study are in favor of it, and in particular they deem that receiving the continuous support of conversational AI technology may improve patients’ adherence to their recommendations. Although the statistical analysis of data collected in this study does not yet show a clear advantage deriving from this integration, the group whose participants received the support of the m-PHA showed a significant positive trend of reduction of symptoms related with obsessivity and compulsivity and positive symptom distress.
Acknowledgments
The research leading to these results has received funding from the European Union’s Horizon 2020 program under grant agreement 826266: COADAPT.

Conflicts of Interest
None declared.

References

https://formative.jmir.org/2021/12/e30053
JMIR Form Res 2021 | vol. 5 | iss. 12 | e30053 | p.17
(page number not for citation purposes)


Abbreviations

AI: artificial intelligence
CBT: cognitive behavioral therapy
GSI: Global Severity Index
MoCA: Montreal Cognitive Assessment
m-PHA: mobile personal health care agent
OSI: Occupational Stress Indicator
PSDI: Positive Symptom Distress Index
PSS: Perceived Stress Scale
PST: Positive Symptom Total
SCL-90-R: Symptom Checklist-90-Revised
SMT: stress management training

https://formative.jmir.org/2021/12/e30053 JMIR Form Res 2021 | vol. 5 | iss. 12 | e30053 | p.18 (page number not for citation purposes)
An Electronic Patient-Reported Outcome Mobile App for Data Collection in Type A Hemophilia: Design and Usability Study

Francesco Petracca1, MSc; Rosaria Tempre2, MSc; Maria Cucciniello1,3, PhD; Oriana Ciani1,4, PhD; Elena Pompeo2, MSc; Luigi Sannino3, MSc; Valeria Lovato2, PhD; Giancarlo Castaman5, MD; Alessandra Ghirardini2, MSc; Rosanna Tarricone1,6, PhD

1Centre for Research in Health and Social Care Management (CERGAS), Government, Health and Non Profit Division, SDA Bocconi, Milan, Italy
2Roche SpA, Monza, Italy
3University of Edinburgh Business School, Edinburgh, United Kingdom
4Institute of Health Research, University of Exeter Medical School, Exeter, United Kingdom
5SODc Malattie Emorragiche e della Coagulazione, Azienda Ospedaliero Universitaria Careggi, Firenze, Italy
6Department of Social and Political Sciences, Bocconi University, Milan, Italy

Corresponding Author:
Francesco Petracca, MSc
Centre for Research in Health and Social Care Management (CERGAS)
Government, Health and Non Profit Division
SDA Bocconi
Via Sarfatti, 10
Milan, 20136
Italy
Phone: 39 02 58365257
Email: francesco.petracca@unibocconi.it

Abstract

Background: There is currently limited evidence on the level and intensity of physical activity in individuals with hemophilia A. Mobile technologies can offer a rigorous and reliable alternative to support data collection processes but they are often associated with poor user retention. The lack of longitudinal continuity in their use can be partly attributed to the insufficient consideration of stakeholder inputs in the development process of mobile apps. Several user-centered models have been proposed to guarantee that a thorough knowledge of the end user needs is considered in the development process of mobile apps.

Objective: The aim of this study is to design and validate an electronic patient-reported outcome mobile app that requires sustained active input by individuals during POWER, an observational study that aims at evaluating the relationship between physical activity levels and bleeding in patients with hemophilia A.

Methods: We adopted a user-centered design and engaged several stakeholders in the development and usability testing of this mobile app. During the concept generation and ideation phase, we organized a need-assessment focus group (FG) with patient representatives to elicit specific design requirements for the end users. We then conducted 2 exploratory FGs to seek additional inputs for the app’s improvement and 2 confirmatory FGs to validate the app and test its usability in the field through the mobile health app usability questionnaire.

Results: The findings from the thematic analysis of the need-assessment FG revealed that there was a demand for sense making, for simplification of app functionalities, for maximizing integration, and for minimizing the feeling of external control. Participants involved in the later stages of the design refinement contributed to improving the design further by upgrading the app’s layout and making the experience with the app more efficient through functions such as chatbots and visual feedback on the number of hours a wearable device had been worn, to ensure that the observed data were actually registered. The end users rated the app highly during the quantitative assessment, with an average mobile health app usability questionnaire score of 5.32 (SD 0.66; range 4.44-6.23) and 6.20 (SD 0.43; range 5.72-6.88) out of 7 in the 2 iterative usability testing cycles.

Conclusions: The results of the usability test indicated a high, growing satisfaction with the electronic patient-reported outcome app. The adoption of a thorough user-centered design process using several types of FGs helped maximize the likelihood of sustained retention of the app’s users and made it fit for data collection of relevant outcomes in the observational POWER study. The continuous use of the app and the actual level of engagement will be evaluated during the ongoing trial.

https://formative.jmir.org/2021/12/e25071  JMIR Form Res 2021 | vol. 5 | iss. 12 | e25071 | p.20 (page number not for citation purposes)
Introduction

Background

Patients with hemophilia A, an X-linked recessive bleeding disorder that occurs in approximately 1 in 5000 live male births, suffer from bleeding episodes, especially in their joints and muscles, and moderate impairment of balance and mobility associated with reduced bone mineral density in both adolescence and adulthood [1]. Because of these limitations, patients with hemophilia typically exhibit reduced levels of physical activity [2].

Few empirical studies have reported on the level and intensity of physical activity in small cohorts of children and adolescents [3-5] and in adult populations with hemophilia A [6-9]. These studies were based on either accelerometers or patient-reported questionnaires and showed that the recommended quantity and quality levels of physical activity were often unmet, with the degree of joint damage accounting for only a small fraction of the observed variability [10]. However, no study has measured the physical activity levels in the hemophilia population and evaluated the correlation between the sequelae of different bleeds and the consequent limitations on physical activity levels.

POWER, a multicenter, noninterventional, prospective study aims at contributing to fill this gap by evaluating the relationship between physical activity levels (and intensity) and bleeding in a target population of approximately 150 individuals aged between 12 and 50 years with severe (Factor VIII <1%) or moderate (Factor VIII ≥1% and Factor VIII ≤2%) hemophilia A without inhibitors against Factor VIII. The study was approved by the Ethics Committee of each participating clinical center and registered at ClinicalTrials.gov (NCT04165135).

This study leverages the widespread availability, low cost, and high degree of reliability of mobile technologies [11] to support the collection of significant amounts of data, including physical activity levels and patient-reported outcome measures (PROMs) [12]. In the field of hemophilia, previous studies have addressed the potential of telehealth-delivered interventions and mobile health (mHealth) solutions to enhance patient adherence to medication and promote independence in disease management [13], improve record keeping [14], and create patient communities that facilitate the interaction of people with hemophilia [15], particularly when they move to adult treatment centers and may report significant feelings of isolation [16]. Although the potential of mHealth technologies to support data collection processes in hemophilia is essentially unexplored, data collection in the POWER study instead relies on 2 different digital devices. The physical activity levels were measured daily in terms of active minutes, the metabolic equivalent of tasks, and the step counts by a wearable fitness device (also called fitness tracker) used continuously during the study participation.

Concurrently, other relevant self-reported outcome domains (bleeds, medications used for bleeds treatment, health-related quality of life, visual analog scale for pain, etc) were registered through an electronic patient-reported outcome (ePRO) app installed on smartphones or tablets after enrollment.

Although a fitness tracker collects physical activity levels through passive sensing, without requiring any extra effort to input data and with very limited engagement with the device besides the need to wear it, the ePRO app requires sustained active input by study participants.

Objectives

User engagement and the continuous app use are persisting challenges in mHealth app implementation [17], with poor user retention being observed in the real world [18,19]. This is also true for ePRO systems, whose interfaces should be continuously monitored and improved to reduce the attrition rates in clinical trials and to enhance the retention postimplementation in clinical practice [20]. A decreasing adherence rate to electronic reporting has also been observed in previous studies that involved digital solutions for patients with hemophilia [21-23].

This effect can be partly attributed to the lack of stakeholder input in the development of mHealth technologies: the apps are often made available to the public without sufficient attention devoted to their design [24] and without a thorough understanding of the context of their proposed deployment and the needs of their end users, regardless of whether they are patients, caregivers, or clinicians [25]. Therefore, to maximize electronic outcome reporting and to ensure continuous data collection throughout the POWER study, we incorporated users’ expectations, experiences, and needs in the design process of the ePRO app. This paper reports on the process adopted for the development of a mobile app aimed at collecting PROMs in patients with hemophilia A during the POWER study.

Methods

Theoretical Framework: User-Centered Design

Approach

We adopted a user-centered design (UCD) and engaged prominent stakeholders in the development of the ePRO app [26]. Among the existing design methods of mHealth apps, UCD primarily focuses on the tasks or activities that the users must accomplish and identifies the corresponding user needs to tackle [27].

According to the UCD approach, during the concept generation phase, a thorough investigation of the needs is conducted to understand the environment of the end users and their requirements. On the basis of this investigation, a set of
functional requirements were identified for the ePRO app through thematic analysis, leading to the design of a prototype. The next phases of the design cycle included continuous, iterative evaluation and refinement of the prototype and its usability testing.

**Study Design**

To improve the app rapidly during the user-centered process, we used focus groups (FGs) as a human factor research technique [28], which could provide appropriate evaluations owing to their flexibility and their ability to probe the participants on key design ideas [29]. Figure 1 shows the various design phases adopted to build the final app.

![Figure 1. Design process of the electronic patient-reported outcome app.](image)

In harmony with UCD, the initial concept generation phase aimed to analyze the environment of the projected end users of the app and to determine their specific requirements in the context of the POWER study. An initial prototype of the app conceived on the basis of the study goals defined in the protocol was used as a starting point and included 3 main screens: (1) the home page, where recent interactions with the app and activities due were listed; (2) the questionnaire page, through which all the PROMs could be accessed and completed; and (3) the My Health Diary page, where all the recorded data could be retrieved.

The need-assessment FG aimed to elicit specific design requirements and specifications for the prototype app and to bring the environment of the intended end users to focus. The script of this initial meeting revolved around: (1) the definition of functional features to be included in the app to maximize the participants’ engagement with data collection; (2) the suggestion of app characteristics, intended as elements that would qualify the app’s visual appearance and its speed or ease of usage; and (3) the discussion of a specific design feature of mHealth studies: the preference between having the app installed on the participant’s own device (thus following a personal device strategy) or receiving a study device at enrollment with the app downloaded on it.

On the basis of the results of this preliminary activity, we implemented a set of functionalities and engagement strategies and discussed them during the following cycles of the design process, when 2 complementary, yet different sets of FGs were run: exploratory FGs (EFGs) and confirmatory FGs (CFGs) [28].

EFGs were conducted to seek additional inputs for improvement, to refine the prototype, and to maximize the likelihood of technology acceptance. A moderator demonstrated the intended use of the mobile app, while participants were asked to provide their feedback on the proposed prototype during an open discussion.

When the finalized version of the app was completed, the CFGs aimed at gathering evidence of its preliminary efficacy. After a brief introduction of the study objectives, the participants were provided with a smartphone or a tablet equipped with the ePRO mobile app, presented with its use case scenarios, and asked to perform a list of 10 different activities (Textbox 1).

![Figure 1. Design process of the electronic patient-reported outcome app.](image)
shown and compared against other commonly used questionnaires that were not strictly designed for mHealth apps [32]. For the app validation process, we presented the participants with the standalone, patient-specific version of the questionnaire, which included 3 main dimensions: (1) ease of use, (2) interface and satisfaction, and (3) usefulness. According to the questionnaire scoring method, higher average values indicated higher usability levels of the app. The MAUQ is yet to be validated in Italian: we used forward and backward translation to adapt the instrument but did not aim to formally validate it, given the small number of panel components.

Textbox 1. Use case scenarios for the electronic patient-reported outcome app (N=10).

| 1. You have just been enrolled in the POWER study. Go to the Study Details section to read the main characteristics of the study and the reasons why your participation is significant. |
| 2. Use the Medication Reminder function to set an alert every 3 days at 6:30 PM. |
| 3. Report a bleeding episode that happened yesterday in your right thigh and was treated with Factor VIII inhibitors. The bleeding ended and resulted in no days away from school or work. |
| 4. Complete the Health Questionnaire you have pending on the home screen. You are feeling great today! |
| 5. Ask the chatbot if your wristband is synchronized with the ePRO mobile device. |
| 6. Fill in the Assigned Treatment questionnaire that you find in the Messages section. |
| 7. After a follow-up with your clinician, your therapeutic plan has been modified. Modify the Medication Reminder settings and set an alert every Tuesday and Thursday at 8:30 AM. |
| 8. The month is coming to an end: fill the Patient and Caregiver Burden questionnaire and report 2 days away from work in the past month. |
| 9. Go back to read the Enrollment Update section where updated statistics on enrolled participants are reported. |
| 10. Check the My Diary section to verify that the information you have inputted are available and easily accessible. |

Participants

We invited different target participants to undergo the various rounds of FGs. The initial need-assessment FG was held in February 2019, with the participation of 4 representatives of patient associations for type A hemophilia. The choice to run this initial meeting with representatives of the most established patient associations in Italy was to maximize the understanding of the particular needs of the targeted end user groups, owing to their continuous and long-term contact with a significant number of affected individuals.

Two different stakeholder groups participated in the subsequent EFGs: 4 hematologists in the clinician FG held in May 2019, whereas 5 patients with hemophilia A in the patient EFG in July 2019.

After the finalization of the refined app, we only involved the intended end users during the CFG phase, with patients and their parents (when patients were minor) being asked to ultimately test the app. Prototype validation through usability testing is usually achieved with 2 to 3 cycles of testing [33], whereas for data saturation, small samples of 5 participants typically identify approximately 80% of the usability issues [34,35]. In this study, 2 different meetings were held in November 2019 and January 2020, with 4 and 5 participants, respectively.

Clinicians participating in the EFG were selected among the study centers on a voluntary basis, whereas patients for both EFGs and CFGs were recruited through the patient associations involved in the initial stage of the design process. Health care professionals were eligible if they were hematologists specializing in bleeding disorders and could speak and read Italian. Eligibility criteria for patients included: confirmed diagnosis of hemophilia A, the ability and willingness to complete outcome questionnaires on the ePRO app, the ability to provide their written consent, and the ability to interact in Italian. Parents of underage patients were invited to participate if their children met the clinical eligibility requirements under the study protocol and if they were willing to provide their informed consent and to test the ePRO app. No rewards or compensation were offered to the individuals participating in the study, whereas a participation fee was awarded to the patient associations involved.

Data Analysis

All FGs occurred in a conference room, lasted for approximately 90 minutes, and were recorded (audio only) and then professionally transcribed. We obtained the written informed consent from all the participants before the start of the meeting. The participants were asked to provide essential personal data, including their age, their professional role (if applicable), and their self-reported comfort in the use of mobile devices. During FGs, we used sticky notes to collect the recommendations of the participants regarding design changes and content review. The analysis of the need-assessment and exploratory FGs was facilitated by Dedoose qualitative software (SocioCultural Research Consultants), a web-based platform for mixed methods analysis [36] that enabled the identification of recurrent themes and the development of a coherent coding index. An inductive analysis was performed by 2 researchers (FP and MC), with emerging themes identified by the systematic reading and coding of the transcripts. Different opinions between the coders were discussed by the research team to reach a consensus.
Results

Participant Characteristics
Of the 22 participants, 11 (50%) were patients, 4 (18%) were patient representatives, 4 (18%) clinicians, and 3 (14%) parents of young patients. Detailed information about the category, age, and gender of the study participants by each research stage has been presented in Table 1.

Table 1. Participant characteristics (N=22).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Need-assessment FG, n (%)</th>
<th>EFGs, n (%)</th>
<th>CFGs, n (%)</th>
<th>Total, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caregiver</td>
<td>0 (0)</td>
<td>1 (11)</td>
<td>6 (67)</td>
<td>3 (33)</td>
</tr>
<tr>
<td>Clinician</td>
<td>0 (0)</td>
<td>4 (44)</td>
<td>0 (0)</td>
<td>4 (18)</td>
</tr>
<tr>
<td>Patient</td>
<td>0 (0)</td>
<td>5 (56)</td>
<td>6 (67)</td>
<td>11 (50)</td>
</tr>
<tr>
<td>Patient representative</td>
<td>4 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>4 (18)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;30</td>
<td>0 (0)</td>
<td>4 (44)</td>
<td>4 (44)</td>
<td>8 (36)</td>
</tr>
<tr>
<td>30-39</td>
<td>0 (0)</td>
<td>1 (11)</td>
<td>1 (11)</td>
<td>2 (9)</td>
</tr>
<tr>
<td>40-49</td>
<td>1 (25)</td>
<td>0 (0)</td>
<td>1 (11)</td>
<td>2 (9)</td>
</tr>
<tr>
<td>50-59</td>
<td>1 (25)</td>
<td>2 (22)</td>
<td>1 (11)</td>
<td>4 (18)</td>
</tr>
<tr>
<td>&gt;60</td>
<td>2 (50)</td>
<td>2 (22)</td>
<td>2 (22)</td>
<td>6 (27)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>3 (75)</td>
<td>6 (67)</td>
<td>7 (78)</td>
<td>16 (73)</td>
</tr>
<tr>
<td>Female</td>
<td>1 (25)</td>
<td>3 (33)</td>
<td>2 (22)</td>
<td>6 (27)</td>
</tr>
</tbody>
</table>

aFG: focus group.
bEFG: exploratory focus group.
cCFG: confirmatory focus group.

Concept Generation: Need-Assessment FG

Thematic analysis of verbatim transcripts of the need-assessment FG led to the identification of 6 themes that were of help in refining the initial prototype.

First, the participants stated that it was paramount to guarantee maximum simplification of the app functionalities, ensuring that the burden on individuals was minimized to what was strictly necessary for study purposes. One patient representative said the following:

Rather than improving the quality and quantity of the app experience, I believe it is fundamental to make it easy, simple, user-friendly, and self-explanatory, making it possible to access it also for somebody who would not want to know how it actually works.

[Patient Representative, 53 years]

Second, the participants made specific recommendations about the sense-making process related to the app, which should leverage the intrinsic motivation of individuals to be part of a noteworthy initiative where the individual benefit is negligible when compared with the profit for the entire community of interest. Therefore, numerous patient representatives suggested the inclusion of specific functionalities to give feedback on the study progress and make every individual part of a larger community. For example, one patient representative vigorously emphasized this point, as follows:

The appeal and the willingness to participate and continuously input data are based on the fact that somebody believes in what the study is proposing, everything else is just a plus...you must necessarily find a way to transfer the sense of what you are asking people to do. [Patient Representative, 64 years]

Third, the FG participants suggested that the app should ensure maximum integration, refraining from any request for data or for doing tasks that can already be performed through different means.

Fourth, a common theme pertained to the need to limit the feeling of external control that a digital solution could exercise on individuals, which can ultimately be linked back to the need to ensure that the participants were reassured about the study’s purposes and modalities. For instance, one participant said the following:

No one should ever get the feeling of being controlled by the app or somebody behind it. I have no need for further sources of control that can make me feel more ill than I already am. [Patient representative, 53 years]

The aesthetics of the user interface were identified as another factor to improve, with the FG participants advising to adopt different color sets to make the layout catchier, but not trivial, than the initial one and to enable individuals to tailor some of the app’s graphical features to their personal preferences.
One final emerging theme pertained to the type of device used during the POWER study. The participants unanimously agreed that providing them with an additional device exclusively for study purposes would prove excessively burdensome and would not help maximize patient engagement, probably generating increased attrition rates instead. One patient representative stated as follows:

_Honestly speaking, to propose in 2019 to participate in a study in which individuals are obliged to use an additional device could be detrimental to the overall study participation._ [Patient Representative, 40 years]

**Changes to the App After Need-Assessment**

Consistent with the demands expressed during the preliminary phase and coded through qualitative analysis, the app underwent the following modifications and additions:

- **The Study details and Enrollment update** pages were added to provide the participants with valuable information about the study and its status, leveraging the required sense-making process;
- Environmental alerts were added to prompt the participants to input the required outcome data;
- Users were given the opportunity to tailor the design features of the app based on their preferences;
- Acknowledgments on activity completion were reinforced to maximize efficiency by providing visual confirmation that the inputted data had been registered;
- A **Treatment reminder** functionality was included in the app and made accessible on a voluntary basis.

**Prototype Refinement: EFGs**

During the 2 EFGs, the clinicians and patients with hemophilia were asked to provide feedback on the refined prototype version of the app to achieve rapid incremental improvements.

During the clinician EFG, the participants univocally acknowledged the significant contribution of the POWER study to provide up-to-date information on the population of interest. Four main codes were identified during the discussion and considered as recommendations to further improve the app.

First, participants suggested avoiding duplications, as some of the sections were potentially overlapping and were not necessarily mutually exclusive. This had primarily related to the structure of the Home page, which was initially designed as a repository of the most recent information included in all the other sections of the app.

Second, a recurrent theme pertained to the graphical interface of the app, against which clinicians suggested facilitating the detection of different domains with corresponding bright colors. One clinician explained the following:

*I find the app look slightly monotonous...to better highlight the various items and domains, different colors could be used across the different sections of the app.* [Clinician, 62 years]

Additional recommendations focused on the study update section. The Study details and Enrollment update sections that were added to the app after the need-assessment phase were appreciated but considered not adequately positioned within the app for accessibility.

Finally, a fourth emerging theme pertained to the possible uses of the app outside of the study settings. To further reduce the patient burden in completing the questionnaires, clinicians suggested machine-readable formats, voice learning, and voice recording. However, these proposals were not technically feasible given the timeframe of the POWER study and, therefore, could not be embedded in the current version of the app.

During the patient EFG that followed, the participants were presented with a refined version of the app that incorporated the main comments collected during the clinician EFG in terms of outlook and content organization.

The coding analysis highlighted several recurrent topics that were brought to attention during the meeting.

First, patients reported their need for support in daily management and coping, expressing a strong willingness to feel a tighter bond with their clinicians and the hope that this could be channeled through enhanced data sharing via the app. In the context of persisting issues in finding appropriate and continuous type of support to sustain patients with hemophilia A daily, there was wide acknowledgment of the potential of digital technologies in closing the existing gap with health care professionals. One patient described this as follows:

_To have somebody to actually follow us through our daily struggle and provide us with prompt feedback would mean a lot...sometimes you just have the impression that you get a pre-set response hours later your request._ [Patient, 18 years]

As a second element, the participants confirmed the need to be supported and facilitated in their participation in the study and in the use of the app to relieve the burden linked to data entry that could cause disaffection with the app, if it were not well supported.

Regarding the interface design, the participants only suggested minor revisions in the layout of some questionnaires, particularly to make the reporting of bleeding events more straightforward. Furthermore, one user recommended substituting the human body where bleeds were indicated to make it less stylized and more realistic.

On the basis of the inputs collected during the EFGs, additional graphical refinements were introduced to further streamline participation in the study: (1) a new chatbot was included to support individuals in finding information related to the app and its functionalities; (2) an additional screen with visual feedback on the number of hours the wearable device had been worn was added to ensure the per-protocol minimum (10 hours/day) was reached and that the observed data were actually registered. The final display of the app home page, chatbot functionality, and bleeding reporting have been shown in **Figure 2**.
Usability Evaluation

During the CFGs, participants tested the final version of the app, followed a list of 10 use case scenarios (Textbox 1), and completed the MAUQ to assess their satisfaction with the system usability.

The average scores showed an upward trend: the overall MAUQ score increased from 5.32 (range 4.44-6.23; SD 0.66) to 6.20 (range 5.72-6.88; SD 0.43) over the 2 CFGs, showing increasingly positive feedback on the app usability as its design was further refined. Considerable improvements were observed especially in terms of perceived usefulness, which increased from 4.80 (SD 0.92) to 6.24 (SD 0.40), and the system information arrangement of the app (from 5.30 to 6.17).

During the first CFG, participants reported poor responsiveness of certain features and a few technical problems, primarily when filling some of the PROMs or when using the chatbot. These issues ultimately affected the reported usefulness and the interface evaluation of the app. The technical difficulties were addressed between the 2 meetings, and the app’s usability ratings improved accordingly. During both the meetings, all participants reported being intuitively able to launch the app and requiring minimal support from the study team during the testing.

On the basis of the inputs provided by the CFG participants through qualitative comments, minor additional changes were included in the finalized version of the app that is currently being used in the POWER study.

Discussion

Principal Findings

We applied UCD for the development of the ePRO app currently in use in the POWER observational study, adopting an iterative process in which progressive modifications were informed not only by the participants’ inputs but also by the technical feasibility of the proposals. This approach has been identified as particularly effective when adapted to mHealth apps [37], with different frameworks being used as powerful alternatives to the shared design based on the end users’ preferences [27]. A recent integrative review analyzed studies that employed a qualitative methodology for the design, development and testing of mHealth apps, identifying 69 articles, and proposing an integrated methodology structured in 4 different sessions [38]. These results confirm the continuous growth in the literature on user-centered approaches for app development. However, all the design studies included in the review aimed to develop apps to support individuals in actively managing their disease through the adoption of behavior change techniques [39] and other cognitive and emotional strategies.

In contrast, none of the studies intended to support the design of data collection apps, with the exception of a single article that focused on the development process of an app for conducting population surveys, but it was meant to be used by health care professionals only [40]. The design process of the ePRO app for the POWER study is a novel attempt to adopt participatory approaches (and UCD, more specifically) to support the collection of patient-reported data. Although we did not aim to elicit improved self-management behaviors, maximized technology acceptance is a fundamental prerequisite to increase the continuity in data collection.
entry and the likelihood of a study’s success. As the precise aims that the app had to pursue were already explicit in the POWER study protocol (ie, to maximize the likelihood of collecting robust outcome data), the design process did not start from scratch and revolved around the how rather than the what.

Despite this difference in the ultimate aim of the app, some of the adopted design strategies we used are comparable with recurring themes in previous studies.

To sustain user engagement, several of the included features and strategies aimed at powering the intrinsic motivation of the participants, defined elsewhere as altruistic motivation [41]. This is coherent with the belief that although strategies that rely on extrinsic motivation only (doing something that can lead to an identifiable outcome) can be effective in the short term, when individuals are intrinsically motivated (doing something which is felt inherently enriching), they tend to achieve better results and guarantee their continued participation in the long run [42].

As a result, increased attention was given to the Study details and the Enrollment update sections, that intend to make the individuals feel they were part of a community and of a mutual effort, which was certified by the growing number of active sites and study participants enrolled.

Concurrently, we included multiple strategies aimed at making the individual experience with the app as efficient as possible, such as the acknowledgment notification to give visual confirmation of the recording of completed questionnaires and the chatbot function to facilitate users in the management of any technical or content-related issues. Furthermore, the questionnaires were tested and revised multiple times to streamline their completion and to minimize the participants’ burden.

Third, to further facilitate the engagement activation process, we tried to take on the challenge of personalization, realizing that technologies should focus on each individual as unique, even in their communication preferences and approaches to data collection processes [43]. Personalization is a recurring theme in app design studies but is typically implemented for ad hoc self-management functionalities or goal setting [44-47]. Instead, given the need to collect the same outcome data for all participants in the POWER study, with no room for content personalization, we included the possibility for users to tailor some of the app graphics and added individual functionalities, such as the treatment reminder, which could be accessed on a voluntary basis. These enhancements do not have a potential interventional effect on the outcomes of interest, and may help generate a greater, personalized bond between the individual user and the mobile app.

In addition to the app content, another domain that might sustain user engagement pertains to the type of mobile device used during the study (either a personal or a study device). Although this ambivalence is a specific feature of mHealth studies, very few studies have attempted to compare the 2 strategies empirically in terms of adherence, with inconclusive statistical results [48]. Despite this debate, there is still no unambiguous settlement on the use of personal devices. However, this strategy, which also goes by the name Bring Your Own Device, has been unquestionably identified as the preferable solution during this app’s development process. This option is not exempt from potential shortcomings associated with the need to develop apps that are compatible with a wide range of systems and security settings [49], the potential selection bias that excludes population segments who do not own a smartphone [50], and the impossibility of locking down the device and maximizing the methodological accuracy. Nevertheless, a personal device presents one major advantage that counterbalances all of the previous shortcomings, as it enhances the convenience for patients, potentially reducing their burden and adopting a pragmatic, real-world approach that replicates a setting more familiar to all the participants.

Finally, to enhance the engagement with the ePRO app, we strengthened the connection with the fitness tracker that records physical activity levels. Although wearables are standard technologies that do not allow for discretion in their design and were previously perceived as highly acceptable by patients with severe hemophilia [7], we reinforced their linkage with the ePRO app by including an ad hoc screen that provided participants with feedback on the number of daily hours the tracker has been actually worn and confirmed whether the per-protocol minimum for physical activity levels to be actually registered (10 hours) was being met.

In addition, the ePRO app may indirectly activate engagement by improving communication with the providers using previously recorded data during routine consultations. The need to be more closely monitored by their physicians was reported by the patients during the FGs and has also emerged in previous app design studies [51,52]. Furthermore, the enhancement of communication links between the health care professionals and the patients, as well as the capture of patient-reported data, are considered among the 4 primary ways through which mHealth can improve the management of hemophilia A [14].

By the end of the design process, we achieved broad agreement that the app was easy to use and had an appealing layout, both preconditions for its sustained use over time. As shown previously by other studies [37], 2 iterative cycles of usability testing were sufficient to reach a satisfactory consensus among the participants and the potential end users. The comparisons of usability with other mobile apps that underwent a thorough development processes are difficult to make because this app is not directly aimed at self-management, and this study was among the first to adopt the MAUQ.

Strengths and Limitations

This research was based on a large sample of participants and included various perspectives by considering the views and needs of the patients, their parents, their representatives, and those of the clinicians during the development process.

Moreover, this study emphasized that user-centered approaches can be applicable, and possibly even more significant, to the development of digital solutions for populations affected with rare diseases, which, by definition, require unique considerations that may be less well-known compared with those for other chronic disease populations. At the same time, the application of the selected methodology to a rare disease group generated additional challenges linked to the difficulties in selecting and
enrolling the participants. Working with smaller sample sizes in rare study populations may be the only way to study them, particularly for methodologies that require the physical presence of participants [53], but several limitations should be acknowledged. First, in terms of patient selection, although we always targeted 5 participants in each of the FGs, in a couple of cases, one of the intended participants communicated their unavailability at short notice because of hemophilia-related issues. Furthermore, to facilitate the identification of the potential participants, we not only adhered to the inclusion criteria identified for the POWER study, but also invited patients with hemophilia A with different severity and inhibitor levels. Although some of the participants were not within the target population of the study, all patients with hemophilia A were subject to similar outcome data collection in standard clinical practice and were, thus, entitled to bring their contribution to the development of the app.

Second, the number of practitioners involved was relatively small and their points of view may not be representative of the entire clinician population or of the organizations they represented.

An additional limitation pertains to usability, which was only evaluated before the field testing of the app in the study and was not assessed longitudinally. Usability evaluation should be a continuous process and should not be limited before dissemination in the field. Additional usability evaluations were planned during the study and at the end of the study through ad hoc use metrics aimed at analyzing the use trajectories. Moreover, we based our usability assessment only on the end user testing and did not properly include any heuristic evaluation involving computer scientists, nor assessed the time to task completion. The mobile app was subject to continuous technical evaluations, and the technical development went hand in hand with the content design.

Finally, the UCD process we adopted was mainly based on user and clinical expert involvement, which are however just 2 of the developmental factors could be included in the design phase of an app [54]. We have noted that alternative methodologies have been developed and they are becoming popular in studies on mobile devices, such as the experience sampling method [55].

Conclusions

The ePRO app will serve as a data collection platform in the POWER observational study. Because all the outcome data collected by the app are directly inputted by the patients, UCD supported the identification of user requirements and the refinement of the app. This process will hopefully meet the users’ expectations and maximize their continuous use of the app throughout the study. The actual level of engagement will be properly monitored during the ongoing POWER study, whereas future research results will assess the effectiveness of this app and demonstrate the value of the development process described here.

Acknowledgments

The authors would like to thank the clinicians, the patients, and the patient representatives who participated in the design process and contributed their ideas to this research. This study was sponsored by Roche SpA. The views expressed in this paper are those of the authors, and no aspects of the study have been omitted.

Conflicts of Interest

FP, MC, OC and RT all reported grants from the European Union’s Horizon 2020 Research and Innovation Program under grant agreement 779306. FP, MC, OC, and RT are also involved in a randomized controlled trial to evaluate a mobile supportive care app for patients with metastatic lung cancer. RT, EP, LS, VL, and AG are employees of Roche SpA. GC is on the advisory boards or is a speaker in company-sponsored symposia for Alexion, Bayer, Sanofi, Roche, Biomarin, Takeda, Novo Nordisk, Werfen, Grifols, Kedrion, LFB, Uniqure, and SOBI.

References


Abbreviations

CFG: confirmatory focus group
EFG: exploratory focus group
ePRO: electronic patient-reported outcome
FG: focus group
MAUQ: mHealth app usability questionnaire
mHealth: mobile health
PROM: patient-reported outcome measure
UCD: user-centered design
Use and Appreciation of a Web-Based, Computer-Tailored Diet and Physical Activity Intervention Based on the Self-determination Theory: Evaluation Study of Process and Predictors

Juul M J Coumans¹, MSc; Anke Oenema²,³, PhD; Catherine A W Bolman¹, PhD; Lilian Lechner¹, PhD

¹Department of Health Psychology, Faculty of Psychology, Open University of the Netherlands, Heerlen, Netherlands
²Department of Health Promotion, Faculty of Health Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands
³Care and Public Health Research Institute, Faculty of Health Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

Corresponding Author:
Juul M J Coumans, MSc
Department of Health Psychology
Faculty of Psychology
Open University of the Netherlands
Valkenburgerweg 177
Heerlen, 6419 AT
Netherlands
Phone: 31 45 576 2635
Email: juul.coumans@ou.nl

Abstract

Background: eHealth is a promising tool for promoting lifestyle behaviors such as a healthy diet and physical activity (PA). However, making people use interventions is a crucial and challenging problem in eHealth. More insight into use patterns and predicting factors is needed to improve future interventions.

Objective: This study aims to examine the use, predictors of use, and appreciation of a web-based, computer-tailored, dietary and PA promotion intervention, MyLifestyleCoach, which is based on the self-determination theory. First, we depict the participants’ flow in the intervention and identify moments when they are likely to discontinue use. Second, we investigate whether demographic, motivational, and program-related characteristics predict the use of several intervention elements. Finally, we report the appreciation scores for the intervention and the participant and program characteristics associated with these scores.

Methods: This study was based on data from web-based self-report questionnaires. Here, objectively assessed intervention use data were analyzed from participants randomized to the intervention condition. Multiple stepwise (logistic) regression analyses were conducted to examine the predictors of intervention use and evaluation scores.

Results: Our findings indicate a low full completion rate for the intervention among those who chose and completed the diet module (49/146, 33.6%), the PA module (2/12, 17%), and both modules (58/273, 21.2%). Several points in the intervention where participants were likely to stop using the intervention were identified. Autonomous and intrinsic motivation toward diet were related to the completion of the initial sessions of the intervention (ie, the opening session in which participants could choose which module to follow and the first session of the diet module). In contrast, controlled motivation was linked to the completion of both modules (initial and follow-up sessions). Appreciation scores were somewhat positive. Appreciation was predicted by several motivational constructs, such as amotivation and basic psychological needs (eg, competence) and program-related features (eg, number of completed sessions).

Conclusions: This study adds meaningful information on the use and appreciation of a web-based, computer-tailored dietary and PA intervention, MyLifestyleCoach. The results indicate that different types of motivations, such as extrinsic and intrinsic motivation, are at play at the points when people are likely to stop using the intervention. The intervention was appreciated fairly well, and several motivational constructs and fulfillment of basic psychological needs were associated with appreciation. Practical implications of these findings have been provided in this study.

(JMIR Form Res 2021;5(12):e22390) doi:10.2196/22390
KEYWORDS
diet; physical activity; eHealth; self-determination theory; motivational interviewing; process evaluation; nonusage attrition

Introduction

Background

Personalized eHealth interventions are promising for promoting a wide array of healthy lifestyle behaviors, such as physical activity (PA) and a healthy diet [1]. The true potential of such an intervention can only be reached when people are sufficiently exposed to its content [2]. However, many people do not use interventions as intended, and many people stop using the intervention before it is fully completed. Eysenbach [3] referred to this phenomenon as nonusage attrition. Research has shown that approximately 50% of the participants used a typical eHealth intervention as intended [4]. There is a general belief regarding the features that make personalized eHealth interventions effective and increase their use. The most essential elements are an increased interaction with a counselor, more frequent intended use, more frequent updates, and more extensive use of dialog support [4]. So far, a detailed understanding is lacking regarding the characteristics of participants who use an intervention as intended, how people navigate through interventions, and where they are likely to stop using the intervention.

To date, several studies have identified predictors of eHealth intervention use. In general, these studies show that age, gender, employment status, a healthier BMI, and lifestyle have been linked to the start, visit and revisit, and use of web-based interventions [5-14]. Mixed results have been found regarding marital status, working status, educational level (although numerous studies show more use for higher-educated people), income, motivation, and self-efficacy as correlates of intervention use [14]. Not only do demographic characteristics and current (lifestyle) behavior explain variance in use but also user engagement, intervention characteristics, and psychological variables could also determine an intervention’s use. Motivation toward a healthier lifestyle could be a crucial factor in use, as it has been related to the initiation and maintenance of health behavior [15,16]. Furthermore, a study found that users who were more autonomously motivated to eat healthily were less likely to stop using the intervention within the first 2 weeks of the program [17]. However, the role of motivation in use has not yet been closely examined [14].

Self-determination theory (SDT) is a macrotheory of human motivation [18]. The SDT postulates that 3 basic psychological needs must be satisfied to maintain optimal performance and well-being. These 3 basic psychological needs are autonomy, competence, and relatedness. When these needs are met, more self-determined forms of motivation are fostered, leading to more engagement in actions to achieve the intended behavior change [18-21]. In the context of use, designing an intervention in which conditions are implemented to support the basic psychological needs may also enhance participation within an intervention. For example, the basic psychological need of autonomy can be implemented in an intervention by giving participants the option to choose what parts of the intervention they want to use and when they want to start with these parts or giving participants a choice on which behavior to work. One study found several characteristics, such as current lifestyle behavior, program features, and amotivation to engage in sufficient PA, to be related to module choice within a multiple health behavior intervention [22].

So far, little is known about use and factors related to use within complex multiple-component (lifestyle) interventions, although this knowledge is very valuable for intervention improvement, particularly concerning use. An intervention that could provide useful information for this purpose is MyLifestyleCoach, a web-based, computer-tailored intervention promoting dietary and PA behavior based on SDT and motivational interviewing. This approach could be one of the underlying mechanisms of intervention use and its effectiveness. In this intervention, people can choose their own way of working through the intervention, that is, which module they want to use (ie, diet, PA, both modules, or no module), and they can decide when to start with the chosen module or modules [23,24]. This approach gives participants autonomy in selecting the behavior they prioritize at a particular moment, which is considered to increase intervention engagement and, ultimately, lower attrition and increase use. Even for people who are already (intrinsically) motivated, this intervention offers tools, such as an action plan, to turn their desire to change behavior into action. In addition to evaluating the use of MyLifestyleCoach, it is important to understand how users appreciate this intervention and whether specific characteristics predict use and appreciation.

Objectives

The first aim of this study is to describe use of the intervention. The second aim is to examine which characteristics are linked to the use of initial and follow-up sessions. The third aim is to examine the appreciation scores for the intervention and what characteristics, especially basic psychological needs, are associated with this appreciation. This study does not shed light on the intervention’s effects; instead, this study provides useful insights for developing future eHealth interventions. For example, it gives a more in-depth understanding of whether providing participants a choice, such as which module to follow and when to follow a module, is beneficial for intervention use.

Methods

Study Design

A 2-group randomized controlled trial (RCT) was conducted in the Netherlands. For this study, observational data of the intervention group of this RCT, called MyLifestyleCoach, was used. Therefore, the control group data were excluded. MyLifestyleCoach is a web-based, computer-tailored intervention that consists of a diet module (I Eat) to promote dietary behavior and a previously tested PA module (I Move) to improve PA levels in Dutch adults. Participants in this intervention could choose which of these modules they would like to take part in both modules, the diet module only, the PA
module only, or no module. Detailed information about the development of the intervention and the design of the RCT, which this study is part of, can be found elsewhere [23,24]. This intervention is theoretically founded on the principles of SDT and uses practical applications of motivational interviewing. This intervention was developed using the intervention mapping protocol [25]. This study was reviewed and approved by the Committee for Ethics and Consent in Research of the Open University of the Netherlands (reference U2018/07266/SVW). This study was registered in the Dutch trial register (NL7333). A data processing agreement with the software developer, who acts in line with the General Data Protection Regulation, has been signed. Furthermore, data that have been exported from the software application are safely stored at the servers of the Open University in accordance with the General Data Protection Regulation.

Participants
The target group for this trial was Dutch adults aged 18-70 years. Participants were recruited using a research panel between October 2018 and May 2019. This research panel sent possible participants an email containing some brief information about the intervention and a link to the intervention website where they could read more information about the goal, procedure, and incentives for the study. The participants’ inclusion criteria were age between 18 and 70 years, an adequate understanding of the Dutch language, and possession of a computer or tablet with access to the internet. Participants who indicated that they had already participated in previous comparable studies of our research group were excluded.

Procedure of the Intervention

Recruitment
A research panel organization sent several emails to recruit participants for this study. In this email, some basic information was provided about this study. Participants could then choose to click on a link leading them to the intervention website with additional information. If participants wanted to participate, they could click on the “I want to participate” button.

Preliminary Assessment and Baseline Questionnaire
First, potential participants had to fill in some questions to assess the previously described inclusion and exclusion criteria of this study and had to sign informed consent. Next, participants were randomly assigned by a computer into the intervention condition or the waiting list control condition (2:1) and filled in the baseline questionnaire. Participants in the intervention condition then continued to the opening session. Participants allocated to the waiting list control condition had no access to the intervention. After the 12-month study period, that is, when they completed the 12-month questionnaire, they were given access to the intervention.

Opening Session
In the opening session, participants were introduced to the program and video coaches. They also received feedback on their dietary and PA behavior using a traffic light system based on the baseline questionnaire results. Participants could receive green advice, indicating that they were already adhering to the guidelines, and following the module was unnecessary; nevertheless, they could have a look at the module. Green advice was provided for diet when they ate at least 2 portions of fruit per day, 250 g of vegetables per day, and fish once a week and consumed no unhealthy snacks per day in line with the Dutch dietary guidelines [26]. For PA, green meant that they were already engaging in ≥150 minutes of moderate to vigorous PA (MVPA) per week according to the Dutch PA guidelines [27]. For diet, orange advice indicated that they were adhering to advice for at least one targeting behavior but not all (eg, consuming sufficient fish but not vegetables). For PA, orange meant that they were engaging in 120-150 minutes of MVPA per week. This cutoff point of 120 minutes of MVPA was chosen based on previous PA guidelines. It was advised to engage in at least 30 minutes of MVPA for at least 5 days per week. Thus, 120 minutes of MVPA (or 4 days of 30-minute MVPA) meant that they almost adhered to the guideline [27], and participants were advised to follow a particular module. Red advice was provided when they did not adhere to any dietary behaviors or had <120 minutes of MVPA per week. Here, participants were strongly advised to follow a particular module. Then, participants could choose whether they wanted to follow the diet or PA module, both modules, or no module. The participants who decided to start with the diet or PA module were given the option to continue to the first session of the module immediately after this opening session or at another moment (within 14 days after the opening session). Participants who decided to follow both modules had to select the module they wanted to start directly and had to choose a date within 14 days after the opening session for the other module. The participants who decided to select no module received an email giving them the option to make a module choice again 2 weeks later. More information can be found elsewhere [22].

Sessions Within Modules
Both the diet and PA modules comprised 4 sessions. In session 1, a healthy diet was explained according to the Netherlands Nutrition Center, or guidelines for sufficient PA levels were provided. Participants were able to see their results on their dietary or PA behavior again. The importance and confidence in eating (more) healthily or engaging in sufficient PA levels were assessed, and feedback was given on this topic. Finally, participants could make an action plan. After 3 weeks from the first session, participants could enter session 2. In this session, they looked back on their perception of the importance of a healthy diet or PA level. They could come up with new reasons to start with the new behavior. Furthermore, they thought about what effects it would have on them if they started with the new behavior (looking forward). Finally, they engaged in a part on coping with difficult situations, including the identification of personal strengths, and could formulate or change the action plan. After 6 weeks from session 1, participants filled in a short questionnaire about their current behavior (diet or PA) and then entered session 3. An assessment took place on participants’ current perception of the importance of a healthy diet or PA and their confidence in achieving this behavior compared with session 1. Participants also received feedback on this assessment. Participants were invited to think back on a problematic situation in which they struggled but managed the achieved behavior.
They received feedback on their current dietary or PA behavior compared with session 1 and could formulate or change the action plan expressed in the previous session of sessions. After 3 months from session 1, they filled in a short questionnaire regarding their current behavior (diet or PA) and then entered session 4, which served as a booster session. Participants could choose several topics from previous sessions that they wanted to do. These topics included feedback on their current behavior compared with session 1, long-term personal motivation and confidence, how to deal with difficult situations, and information on how to maintain their new behavior after the end of the program. Figure 1 shows an overview of the content of these sessions. More detailed information on these sessions can be found in the protocol papers [23,24].

Figure 1. Overview of the content of the sessions in the intervention. PA: physical activity.
Follow-up Questionnaire

After 6 months from when participants completed the baseline questionnaire, both in the intervention and control conditions, they were sent an invitation email to complete the 6-month follow-up questionnaire. Email reminders were sent every week for 4 weeks in total. Participants who completed all questionnaires were entered into a draw for 2 tablets and gift vouchers of up to €50 (US $57.23).

Measurements

The baseline questionnaire assessed demographic characteristics, dietary and PA behaviors, and psychosocial constructs. All these measurements were self-reported.

Demographics

Demographic characteristics included age, gender, education, work status, physical impairment, marital status, weight and height, and health status using a thermometer-style visual analog scale ranging from 0-100. These factors served as control variables in our analysis.

Motivation

Of the psychological constructs measured in this study, only motivation was included. Motivation was assessed with 2 Treatment Self-Regulation Questionnaires, one for dietary behavior and the other for PA behavior [28]. Participants had to indicate the degree to which they agreed with each of the 15 statements on a 7-point Likert scale. There were 3 subscales: amotivation (3 items), controlled motivation (6 items), and autonomous motivation (6 items). This questionnaire did not assess the intrinsic motivation for these health-related behaviors. For that purpose, we included the intrinsic regulation subscale (4 items) from the Dutch Behavioral Regulation in Exercise Questionnaire-2 to determine the intrinsic motivation for PA behavior and an adapted version of the Behavioral Regulation in Exercise Questionnaire-2 to determine the intrinsic motivation for dietary behavior [29]. Participants had to indicate the degree to which they agreed with each of the 4 statements on a 5-point Likert scale. The mean score was calculated for each motivational construct.

Dietary and PA Behavior

Dietary behavior was assessed using a validated Food Frequency Questionnaire. The Food Frequency Questionnaire was extended with questions regarding the size of vegetable and fruit portions based on the study by Huybrechts et al [30]. The outcomes were fruit intake, vegetable intake, fish consumption, and daily consumption frequency of unhealthy snacks. For the calculation of the consumption frequency of unhealthy snacks, we referred to the study by Coumans et al [31]. PA behavior for a typical week in the past month was assessed using the validated Dutch Short Questionnaire to Assess Health [32]. PA behavior was operationalized as the total number of minutes of MVPA by multiplying the frequency (days per week) and duration (hours and minutes per day) of leisure and transport walking, leisure and transport cycling, occupational activities, household activities, gardening, odd jobs, and sports performed with moderate or vigorous intensity.

Process Evaluation

To assess appreciation, participants were asked to give an appreciation score for the whole program on a 10-point scale at 6 months from baseline. People also had to provide a rating for the diet and PA module, which ranged on a scale from 1 (very low) to 10 (very high), if they had completed at least 1 session of the particular module. Furthermore, participants were asked to what extent the program met their basic psychological needs during the intervention on a 5-point Likert scale from 1 (fully disagree) to 5 (fully agree) [33]. A total of 2 items assessed autonomy: (1) participants were asked if they could determine which goals they could set and (2) which information and pieces of advice they could read in the intervention. Relatedness was assessed by 3 items: (1) participants were asked if they felt involved in the intervention, (2) if the intervention was personal, and (3) if they felt supported by the intervention. Competence was assessed with 1 item: participants were asked whether they had confidence in eating (more) healthily and/or engaging in more or sufficient PA. The mean score for each of these basic psychological needs was calculated.

Finally, data on the completeness of sessions were used to determine how many participants used a specific part of the intervention (use). A completeness variable (1=completed and 0=not completed) was created for each session: the opening session and module sessions 1, 2, 3, and 4 of the diet and PA modules of the intervention. When participants finished a session, the completeness variable was set to 1. For this study, 5 use variables were created: (1) finished opening session (1=yes and 0=no), (2) finished the first session of the diet module when only the diet module was chosen (1=yes and 0=no), (3) finished the first sessions of both the diet and PA modules when both modules were chosen (1=yes and 0=no; based on 4 complete sessions), and (5) finished the diet and PA module when both modules were chosen (1=yes and 0=no; based on 8 complete sessions).

Statistical Analysis

Descriptive statistics (mean and SD values) and frequencies (and percentages) were used to depict the characteristics of the participants, the overall flow through the intervention, and appreciation scores. Logistic regression analyses were conducted to examine which personal characteristics (age, gender, education, marital status, work, physical impairment, health status, and BMI) and motivational characteristics were related to use. Use was subdivided into 3 parts according to different points in the intervention: (1) completion of the opening session; (2) the initial module’s session, that is, completion of the first session of the diet module when only the diet module was chosen or the first session of the diet and PA module when both modules were chosen; and (3) the follow-up sessions, that is, completion of all 4 sessions of the diet module when only the diet module was chosen or completion of all 8 sessions of the diet and PA module when both modules were chosen. Furthermore, linear regression analyses were performed to investigate which demographic factors, motivational constructs, and program features were associated with the intervention appreciation scores (overall intervention, diet, and PA module).
All statistical analyses were performed using the statistical software R (version 3.6.0; R Foundation for Statistical Computing). For all regression analyses, a stepwise approach was used in which the demographic variables were entered in the first step, motivational constructs were introduced in the second step, and program features were added in the third step. Variance inflation factors were inspected before conducting the analyses. Statistical significance was set at $P<.05$.

Results

Participants’ Characteristics
The mean age of the sample was 51.9 (SD 13.1) years; there were slightly more women than men participating in the study; 70.3% (545/775) of the sample was highly educated; and 64% (496/775) were employed. The mean BMI of this sample was considered to be slightly overweight. However, the proportion of participants with a healthy weight was the largest. More characteristics are presented in Table 1.
Table 1. Demographic characteristics of the full sample (N=775).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>51.9 (13.1)</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>475 (61.3)</td>
</tr>
<tr>
<td>Men</td>
<td>300 (38.7)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>29 (3.7)</td>
</tr>
<tr>
<td>Medium</td>
<td>201 (25.9)</td>
</tr>
<tr>
<td>High</td>
<td>545 (70.3)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Partner</td>
<td>529 (68.3)</td>
</tr>
<tr>
<td>Single</td>
<td>246 (31.7)</td>
</tr>
<tr>
<td><strong>Work, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>496 (64.0)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>279 (36.0)</td>
</tr>
<tr>
<td><strong>Physical impairment, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>740 (95.5)</td>
</tr>
<tr>
<td>Yes</td>
<td>35 (4.5)</td>
</tr>
<tr>
<td><strong>BMI status</strong>, mean (SD)**</td>
<td></td>
</tr>
<tr>
<td>Underweight</td>
<td>15 (1.9)</td>
</tr>
<tr>
<td>Normal</td>
<td>328 (42.3)</td>
</tr>
<tr>
<td>Overweight</td>
<td>279 (36.0)</td>
</tr>
<tr>
<td>Obese</td>
<td>153 (19.7)</td>
</tr>
<tr>
<td>BMI (kg/m$^2$), mean (SD)</td>
<td>26.5 (5.2)</td>
</tr>
<tr>
<td>Health status (0-100), mean (SD)</td>
<td>69.9 (15.6)</td>
</tr>
<tr>
<td>Amotivation diet (1-7), mean (SD)</td>
<td>2.3 (1.2)</td>
</tr>
<tr>
<td>Controlled motivation diet (1-7), mean (SD)</td>
<td>2.8 (1.2)</td>
</tr>
<tr>
<td>Autonomous motivation diet (1-7), mean (SD)</td>
<td>5.5 (1.2)</td>
</tr>
<tr>
<td>Intrinsic motivation diet (1-5), mean (SD)</td>
<td>3.5 (1.0)</td>
</tr>
<tr>
<td>Amotivation PA$^b$ (1-7), mean (SD)</td>
<td>2.2 (1.3)</td>
</tr>
<tr>
<td>Controlled motivation PA (1-7), mean (SD)</td>
<td>2.7 (1.2)</td>
</tr>
<tr>
<td>Autonomous motivation PA (1-7), mean (SD)</td>
<td>5.6 (1.2)</td>
</tr>
<tr>
<td>Intrinsic motivation PA (1-5), mean (SD)</td>
<td>3.8 (1.1)</td>
</tr>
<tr>
<td>Fruit, mean (SD)</td>
<td>1.4 (1.1)</td>
</tr>
<tr>
<td>Vegetables, mean (SD)</td>
<td>143.1 (80.7)</td>
</tr>
<tr>
<td>Fish (0-7), mean (SD)</td>
<td>1.1 (1.1)</td>
</tr>
<tr>
<td>Unhealthy snacks, mean (SD)</td>
<td>1.5 (1.9)</td>
</tr>
<tr>
<td>MVPA$^c$, mean (SD)</td>
<td>992.7 (836.8)</td>
</tr>
</tbody>
</table>

$^a$Underweight: a BMI value of <18.5 kg/m$^2$; normal weight: a BMI value ranging from 18.5 kg/m$^2$ to <25.0 kg/m$^2$; overweight: a BMI value ranging from 25.0 kg/m$^2$ to <30.0 kg/m$^2$; and obese: a BMI value of ≥30.0 kg/m$^2$.

$^b$PA: physical activity.

$^c$MVPA: moderate to vigorous physical activity.
Description of the Participants’ Flow and Module Use

Figure 2 illustrates the flow of participants. The boxes and text in gray represent the control conditions. This study focuses solely on the use and appreciation of the intervention. Therefore, the control condition was not included, as they did not take part in the intervention in this timeframe. In total, 9806 individuals were directly contacted via the research panel organization. Of these, 23.64% (2318/9806) of individuals visited the study website and clicked on the “I want to participate” button; 16.55% (1623/9806) of these individuals passed the inclusion criteria and signed the informed consent and were randomized into the 2 conditions. Several individuals in the intervention condition did not complete the baseline questionnaire after randomization (315/1090, 28.9%). Of the 775 participants in the intervention condition, 619 (79.9%) made a choice on which module (diet, PA, both, or none) to follow, and 579 (74.7%) participants completed the entire opening session.

Of the 158 participants who chose to follow the diet module only, 8 (5.1%) participants did not choose whether they wanted to start immediately or later or did not fill in when to start with the first session, and 4 (2.5%) participants did not receive an invitation mail for the first session. Of the remaining 146 participants, half of them started immediately, whereas the other half wanted to start later. Of the 73 participants who decided to start immediately, 44 (60%) completed session 1, 32 (44%) completed session 2, 32 (44%) completed session 3, and 35 (48%) completed session 4. Of the 73 participants who decided to start later, 39 (53%) completed session 1, 26 (37%) completed session 2, 28 (38%) completed session 3, and 28 (38%) completed session 4. Approximately 33.6% (49/146) of participants completed all 4 sessions in the diet module.

Within the PA module (n=12, as 2 participants did not receive an invitation mail for the first session), half of the participants decided to start immediately, and the other half wanted to start later. Of the 6 participants who decided to start immediately, 4 (67%) completed session 1, 2 (33%) completed session 2, 1 (17%) completed session 3, and 1 (17%) completed the fourth session. Of the 6 participants who decided to start later, only 1 (17%) completed all 4 sessions. Participants who only chose the diet module were asked after every session, except for the last one, whether they were interested in starting with the diet module. Of the 12 participants who only chose the PA module, 1 (8%) participant was interested in the diet module; however, this person did not complete any session of the diet module.

Of the 339 who chose both modules, 32 (9.4%) did not complete the opening session, and 34 (10%) did not receive an invitation mail for the second module because of a technical error. Most participants decided to start with the diet module (244/273, 89.4%) compared with the PA module (29/273, 10.6%). Full completion rates, that is, those who completed all sessions as intended, can be found in Figure 3. In the right panel, 2 lines are added that represent the people who chose both modules who completed the sessions of the separate diet module (dark gray) or completed the sessions of the separate physical activity module (light gray). Approximately 21.2% (58/273) of participants completed all sessions of both modules. Approximately 5% (4/77) of participants who preferred to start later with the module did not receive an invitation email for the first session of the diet module because of a technical error. Moreover, approximately 25% (2/8) of participants who preferred to start later with the module did not receive an invitation email for the first session of the PA module because of a technical error.

Participants who initially chose no module in the intervention (110/775, 14.2%) were sent an email directing them to the website where they could change their choice. Of the 110 participants, only 4 (3.6%) reconsidered their choice, and 2 (1.8%) chose a module. Of the 110 participants, 1 (0.9%) chose the diet module, 1 (0.9%) chose the PA module (and did not finish any modules), and 2 (1.8%) chose the no module option again.

Finally, 45% (349/775) of participants completed the follow-up questionnaire at 6 months from baseline. Of the 158 participants who chose the diet module, 78 (49.4%) completed the follow-up questionnaire. Of the 14 participants who chose the PA module, 9 (64.3%) completed the follow-up questionnaire. Of the 339 participants who chose both modules, 151 (44.5%) completed the follow-up questionnaire. Of the 108 participants who did not choose any module in the opening session, 58 (53.7%) completed the follow-up questionnaire. Of the 156 participants who did not enter the opening session or made no module choice, 53 (33.9%) completed the follow-up questionnaire. This follow-up measurement included the process evaluation questions.
Figure 2. Participants' flow in the randomized controlled trial. Participants did not need to complete the second session to be able to continue the third session. Therefore, use rates do not necessarily represent a funnel shape. PA: physical activity.
As described in the previous section and as can be seen in the participants' flowchart (Figure 2) and the completion rates in Figure 3, there are several moments within the intervention at which participants stop using the intervention. First, several people did not complete the baseline questionnaire after randomization (315/1090, 28.9%; Figure 2). As we did not have the demographic characteristics of this group, it was not possible to further examine predictors of why they stopped using the intervention. Second, there was a significant number of participants who did not complete the opening session after completing the baseline questionnaire (Figure 2). Third, another group did not start or end the first session of their chosen module (Figure 3). Fourth, numerous people did not complete the sessions as intended, that is, about half of the participants completed the whole intervention once started (Figure 3). Here, we investigate whether there are characteristics associated with use for the latter 3 moments.

The logistic regression analysis (Table 2) showed that participants with a partner and those who had a higher intrinsic motivation to eat healthily were less likely to complete the opening session. Participants with higher scores on autonomous motivation to eat healthily and those with a higher score on intrinsic motivation to engage in sufficient PA were more likely to complete the opening session. Therefore, intrinsic motivation toward PA increased the likelihood of completing the opening session, whereas intrinsic motivation toward a healthy diet decreased the likelihood of completing the opening session.

The logistic regression analyses relating demographic characteristics, motivational constructs, and intervention features with completing the entire intervention are presented in Table 4. The whole intervention could concern the 4 sessions of (1) the diet module, (2) the PA module—owing to the low number of participants that only chose the PA module, predictors of use for the PA module were not further investigated, as it would have been statistically invalid—and (3) both the diet and PA module. For participants who only chose the diet module, it was found that those who received red advice for diet in the opening session compared with orange advice were more likely to complete all sessions within the diet module. For participants who chose both modules, the results showed that older participants and those with a higher controlled motivation toward PA were more likely to complete all sessions of both modules. On the other hand, participants with a higher BMI or more intrinsic motivation to eat healthily were less likely to complete the first session of the diet module. For participants who chose both modules, the results showed that participants having a higher self-reported health status, more controlled motivation to PA, and receiving red advice for PA compared with orange advice were more likely to complete the first sessions of both modules. On the contrary, participants with a partner or those who preferred to start with the PA module were less likely to complete the first sessions of both modules. The results are presented in Table 3. Owing to the low number of participants that only chose the PA module, predictors of use for the PA module were not further investigated, as it would have been statistically invalid.

Finally, the regression analyses relating demographic characteristics, motivational constructs, and intervention features with completing the entire intervention are presented in Table 4. The whole intervention could concern the 4 sessions of (1) the diet module, (2) the PA module—owing to the low number of participants that only chose the PA module, predictors of use for the PA module were not further investigated, as it would have been statistically invalid—and (3) both the diet and PA module. For participants who only chose the diet module, it was found that those who received red advice for diet in the opening session compared with orange advice were more likely to complete all sessions within the diet module. For participants who chose both modules, the results showed that older participants and those with a higher controlled motivation toward PA were more likely to complete all sessions of both modules. On the other hand, participants with a partner or those who preferred to start with the PA module were less likely to complete all sessions of both modules. The full basic models can be found in Multimedia Appendix 1.
Table 2. Results of the stepwise logistic regression analyses (full model) showing variables associated with completing the opening session (N=775)\(^a\).

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Completed opening session (1=yes and 0=no)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds ratio (95% CI; SE)</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>1.31 (0.21-8.09; 0.93)</td>
<td>.77</td>
</tr>
<tr>
<td>Age</td>
<td>1.003 (0.99-1.02; 0.01)</td>
<td>.69</td>
</tr>
<tr>
<td>Gender(^b)</td>
<td>1.27 (0.88-1.84; 0.19)</td>
<td>.20</td>
</tr>
<tr>
<td>Education high(^c)</td>
<td>0.98 (0.66-1.47; 0.21)</td>
<td>.93</td>
</tr>
<tr>
<td>Education low(^c)</td>
<td>1.09 (0.40-2.94; 0.51)</td>
<td>.86</td>
</tr>
<tr>
<td>Marital status partner(^d)</td>
<td>0.60 (0.41-0.89; 0.20)</td>
<td>.01(^e)</td>
</tr>
<tr>
<td>Work employed(^f)</td>
<td>0.97 (0.67-1.41; 0.19)</td>
<td>.87</td>
</tr>
<tr>
<td>Impairment(^g)</td>
<td>1.74 (0.66-4.57; 0.49)</td>
<td>.26</td>
</tr>
<tr>
<td>BMI</td>
<td>1.04 (0.996-1.08; 0.02)</td>
<td>.08</td>
</tr>
<tr>
<td>Health status</td>
<td>1.0005 (0.99-1.01; 0.01)</td>
<td>.94</td>
</tr>
<tr>
<td>Amotivation diet</td>
<td>1.09 (0.88-1.33; 0.10)</td>
<td>.43</td>
</tr>
<tr>
<td>Amotivation PA(^b)</td>
<td>0.90 (0.74-1.10; 0.10)</td>
<td>.30</td>
</tr>
<tr>
<td>Controlled motivation diet</td>
<td>0.95 (0.75-1.20; 0.12)</td>
<td>.68</td>
</tr>
<tr>
<td>Controlled motivation PA</td>
<td>1.12 (0.89-1.42; 0.12)</td>
<td>.33</td>
</tr>
<tr>
<td>Autonomous motivation diet</td>
<td>1.35 (1.05-1.74; 0.13)</td>
<td>.02(^e)</td>
</tr>
<tr>
<td>Autonomous motivation PA</td>
<td>0.80 (0.61-1.04; 0.13)</td>
<td>.10</td>
</tr>
<tr>
<td>Intrinsic motivation diet</td>
<td>0.60 (0.48-0.76; 0.12)</td>
<td>&lt;.001(^f)</td>
</tr>
<tr>
<td>Intrinsic motivation PA</td>
<td>1.38 (1.13-1.68; 0.10)</td>
<td>.001(^f)</td>
</tr>
</tbody>
</table>

\(^a\)The results’ interpretations are reported when all other predictors are held constant. Explained variance $R^2_{\text{tjur}}=0.054; \text{Akaike information criterion}=869.59.$

\(^b\)Female is the reference category.

\(^c\)Medium education is the reference category.

\(^d\)Single is the reference category.

\(^e\)Values represent statistical significance.

\(^f\)Being unemployed is the reference category.

\(^g\)No physical impairment is the reference category.

\(^h\)PA: physical activity.
### Table 3. Results of the stepwise logistic regression analyses (full model) showing variables associated with completing the first session for the diet module and both modules.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Session 1 (1=yes and 0=no)</th>
<th>Both modules&lt;sup&gt;c&lt;/sup&gt;</th>
<th>P value</th>
<th>OR (95% CI; SE)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diet module&lt;sup&gt;a,b&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>2.88 (0.04-221.08; 2.22)</td>
<td>.63</td>
<td>0.06 (0.002-2.39; 1.85)</td>
<td>.14</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.02 (0.99-1.06; 0.02)</td>
<td>.26</td>
<td>1.02 (0.99-1.04; 0.01)</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>Gender&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.69 (0.31-1.56; 0.41)</td>
<td>.38</td>
<td>0.70 (0.38-1.31; 0.32)</td>
<td>.27</td>
<td></td>
</tr>
<tr>
<td>Education high&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1.02 (0.38-2.73; 0.50)</td>
<td>.96</td>
<td>1.39 (0.73-2.66; 0.33)</td>
<td>.32</td>
<td></td>
</tr>
<tr>
<td>Education low&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1.53 (0.19-12.44; 1.07)</td>
<td>.69</td>
<td>1.21 (0.25-5.91; 0.81)</td>
<td>.82</td>
<td></td>
</tr>
<tr>
<td>Marital status partner&lt;sup&gt;g&lt;/sup&gt;</td>
<td>1.03 (0.43-2.45; 0.44)</td>
<td>.95</td>
<td>0.52 (0.28-0.95; 0.31)</td>
<td>.03&lt;sup&gt;h&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Work employed&lt;sup&gt;i&lt;/sup&gt;</td>
<td>1.26 (0.53-3.01; 0.44)</td>
<td>.60</td>
<td>0.84 (0.46-1.54; 0.31)</td>
<td>.57</td>
<td></td>
</tr>
<tr>
<td>Impairment&lt;sup&gt;j&lt;/sup&gt;</td>
<td>9.05 (1.06-77.10; 1.09)</td>
<td>.04&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3.04 (0.79-11.75; 0.69)</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>0.87 (0.79-0.97; 0.05)</td>
<td>.09&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.98 (0.93-1.04; 0.03)</td>
<td>.55</td>
<td></td>
</tr>
<tr>
<td>Health status</td>
<td>1.02 (0.99-1.05; 0.02)</td>
<td>.25</td>
<td>1.04 (1.01-1.06; 0.01)</td>
<td>.001&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Amotivation diet</td>
<td>0.83 (0.49-1.39; 0.26)</td>
<td>.47</td>
<td>1.19 (0.84-1.67; 0.17)</td>
<td>.33</td>
<td></td>
</tr>
<tr>
<td>Amotivation PA&lt;sup&gt;k&lt;/sup&gt;</td>
<td>1.50 (0.95-2.37; 0.23)</td>
<td>.08</td>
<td>0.72 (0.49-1.06; 0.20)</td>
<td>.10</td>
<td></td>
</tr>
<tr>
<td>Controlled motivation diet</td>
<td>1.24 (0.68-2.25; 0.30)</td>
<td>.48</td>
<td>0.70 (0.47-1.06; 0.21)</td>
<td>.09</td>
<td></td>
</tr>
<tr>
<td>Controlled motivation PA</td>
<td>0.65 (0.37-1.15; 0.29)</td>
<td>.14</td>
<td>1.70 (1.13-2.57; 0.21)</td>
<td>.01&lt;sup&gt;h&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Autonomous motivation diet</td>
<td>2.27 (1.17-4.41; 0.34)</td>
<td>.02&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.995 (0.64-1.55; 0.23)</td>
<td>.98</td>
<td></td>
</tr>
<tr>
<td>Autonomous motivation PA</td>
<td>0.69 (0.36-1.33; 0.33)</td>
<td>.27</td>
<td>0.97 (0.61-1.55; 0.24)</td>
<td>.90</td>
<td></td>
</tr>
<tr>
<td>Intrinsic motivation diet</td>
<td>0.53 (0.31-0.91; 0.27)</td>
<td>.02&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.94 (0.66-1.33; 0.18)</td>
<td>.71</td>
<td></td>
</tr>
<tr>
<td>Intrinsic motivation PA</td>
<td>1.05 (0.65-1.33; 0.25)</td>
<td>.83</td>
<td>0.80 (0.58-1.11; 0.17)</td>
<td>.19</td>
<td></td>
</tr>
<tr>
<td>Diet advice green&lt;sup&gt;l,m&lt;/sup&gt; —</td>
<td>—</td>
<td>0.00 (974.90)</td>
<td>.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diet advice red&lt;sup&gt;i&lt;/sup&gt;</td>
<td>2.77 (0.98-7.86; 0.53)</td>
<td>.06</td>
<td>0.59 (0.31-1.12; 0.33)</td>
<td>.11</td>
<td></td>
</tr>
<tr>
<td>PA advice green&lt;sup&gt;i&lt;/sup&gt; —</td>
<td>—</td>
<td>3.18 (0.40-24.92; 1.05)</td>
<td>.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA advice red&lt;sup&gt;i&lt;/sup&gt; —</td>
<td>—</td>
<td>16.82 (1.18-221.17; 1.31)</td>
<td>.03&lt;sup&gt;h&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Module start (later&lt;sup&gt;o&lt;/sup&gt;)</td>
<td>0.55 (0.25-1.22; 0.40)</td>
<td>.14</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>First module (PA&lt;sup&gt;p&lt;/sup&gt;) —</td>
<td>—</td>
<td>0.21 (0.07-0.63; 0.56)</td>
<td>.005&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Physical activity advice was not included, as 3 participants did not receive green advice. Odds ratios were unreliable when this variable was included in the analyses. The results’ interpretations are reported when all other predictors are held constant.

<sup>b</sup>Observations=146; R<sup>2</sup> tjur=0.158; Akaike information criterion=215.19.

<sup>c</sup>Observations=273; R<sup>2</sup> tjur=0.167; Akaike information criterion=366.99.

<sup>d</sup>OR: odds ratio.

<sup>e</sup>Female is the reference category.

<sup>f</sup>Medium education is the reference category.

<sup>g</sup>Single is the reference category.

<sup>h</sup>Values represent statistical significance.

<sup>i</sup>Being unemployed is the reference category.

<sup>j</sup>No physical impairment is the reference category.

<sup>k</sup>PA: physical activity.

<sup>l</sup>Orange advice is the reference category.

<sup>m</sup>Only 2 participants received green advice. Consequently, the odds ratio and SE are less reliable, and CI is not reported.
These variables were not included in the model.

Directly starting with the first session was the reference category.

Choosing the diet module to start with when both modules were chosen was the reference category.
Table 4. Results of the stepwise logistic regression analyses (full model) showing variables associated with completing all sessions when the diet or both modules were chosen.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>All sessions (1=yes and 0=no)</th>
<th>Diet module</th>
<th>Both modules</th>
<th>P value</th>
<th>Both modules</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95% CI; SE)</td>
<td>P value</td>
<td>OR (95% CI; SE)</td>
<td>P value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.05 (0.001-3.91; 2.25)</td>
<td>.18</td>
<td>0.22 (0.003-14.94; 2.16)</td>
<td>.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.04 (0.998-1.08; 0.02)</td>
<td>.07</td>
<td>1.03 (1.002-1.07; 0.02)</td>
<td>.04e</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>0.99 (0.44-2.22; 0.41)</td>
<td>.98</td>
<td>1.08 (0.52-2.25; 0.37)</td>
<td>.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education high</td>
<td>1.55 (0.60-4.04; 0.49)</td>
<td>.37</td>
<td>1.34 (0.61-2.96; 0.40)</td>
<td>.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education low</td>
<td>1.32 (0.17-10.57; 1.06)</td>
<td>.79</td>
<td>1.12 (0.15-8.10; 1.01)</td>
<td>.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marital status partner</td>
<td>0.91 (0.38-2.18; 0.44)</td>
<td>.84</td>
<td>0.51 (0.25-1.04; 0.37)</td>
<td>.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work employed</td>
<td>0.85 (0.35-2.07; 0.45)</td>
<td>.72</td>
<td>0.73 (0.36-1.49; 0.36)</td>
<td>.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impairment</td>
<td>0.88 (0.14-5.32; 0.92)</td>
<td>.89</td>
<td>1.82 (0.36-9.13; 0.82)</td>
<td>.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>0.95 (0.86-1.05; 0.05)</td>
<td>.32</td>
<td>0.95 (0.89-1.02; 0.03)</td>
<td>.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health status</td>
<td>1.01 (0.98-1.04; 0.02)</td>
<td>.63</td>
<td>1.02 (0.99-1.04; 0.01)</td>
<td>.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amotivation diet</td>
<td>1.19 (0.71-1.99; 0.26)</td>
<td>.51</td>
<td>1.27 (0.85-1.89; 0.20)</td>
<td>.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amotivation PA</td>
<td>1.19 (0.76-1.85; 0.23)</td>
<td>.45</td>
<td>0.95 (0.60-1.49; 0.23)</td>
<td>.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controlled motivation diet</td>
<td>1.09 (0.60-1.97; 0.30)</td>
<td>.79</td>
<td>0.43 (0.25-0.75; 0.28)</td>
<td>.003e</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controlled motivation PA</td>
<td>0.63 (0.35-1.13; 0.30)</td>
<td>.12</td>
<td>2.60 (1.48-4.59; 0.29)</td>
<td>.001e</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autonomous motivation diet</td>
<td>1.55 (0.82-2.95; 0.33)</td>
<td>.18</td>
<td>0.92 (0.52-1.53; 0.25)</td>
<td>.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autonomous motivation PA</td>
<td>0.82 (0.44-1.53; 0.32)</td>
<td>.52</td>
<td>0.89 (0.52-1.53; 0.28)</td>
<td>.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intrinsic motivation Diet</td>
<td>0.96 (0.56-1.64; 0.28)</td>
<td>.88</td>
<td>0.83 (0.53-1.30; 0.23)</td>
<td>.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intrinsic motivation PA</td>
<td>1.07 (0.65-1.75; 0.25)</td>
<td>.79</td>
<td>0.92 (0.62-1.39; 0.21)</td>
<td>.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diet advice green</td>
<td>—</td>
<td>—</td>
<td>0.00 (951.69)</td>
<td>.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diet advice red</td>
<td>3.17 (1.11-9.04; 0.53)</td>
<td>.03e</td>
<td>0.39 (0.16-0.90; 0.43)</td>
<td>.03e</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA advice green</td>
<td>—</td>
<td>—</td>
<td>1.69 (0.15-19.40; 1.24)</td>
<td>.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA advice red</td>
<td>—</td>
<td>—</td>
<td>9.89 (0.50-194.10; 1.52)</td>
<td>.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Start (later)</td>
<td>0.56 (0.26-1.23; 0.40)</td>
<td>.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First module (PA³)</td>
<td>—</td>
<td>—</td>
<td>0.05 (0.01-0.50; 1.16)</td>
<td>.01e</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

*aThe results’ interpretations are reported when all other predictors are held constant.

*bObservations=146; R² (tjur)=0.099; Akaike information criterion=211.14.

*cObservations=273; R² (tjur)=0.188; Akaike information criterion=276.69.

*dOR: odds ratio.

*eValues represent statistical significance.

*fFemale is the reference category.

*Medium education is the reference category.

*hSingle is the reference category.

*iBeing unemployed is the reference category.

+jNo physical impairment is the reference category.

+kPA: physical activity.

+lOrange advice is the reference category.

+mOnly 2 participants received green advice. Consequently, the odds ratio and SE are less reliable, and CI is not reported.

*nThese variables were not included in the model.
Appreciation and Its Predictors

After 6 months from baseline, participants were asked to complete the follow-up questionnaire, including the process evaluation measures. These process evaluation measurements focused on the extent to which the program met the participants’ basic psychological needs, which were operationalized as the ratings they gave for autonomy, competence, and relatedness. The mean appreciation score for the intervention as a whole was 6.9 (SD 1.7). Approximately 83.3% (245/294) of participants provided a rating of 6 out of 10 or higher (sufficient). Overall, the mean scores of the process evaluation variables represented neutral (relatedness and competence) to positive scores (autonomy) of the intervention. For autonomy, the average rating was 3.9 (SD 0.8; 291/294, 99%) out of 5. For relatedness, the average rating was 3.1 (SD 0.9; 291/294, 99%) out of 5. For competence, the average rating was 3.0 (SD 1.0; 291/294, 99%) out of 5. The appreciation of the diet module was, on average, 7.1 (SD 1.7; 159/294, 54.1%), whereas the appreciation of the PA module was, on average, 7.4 (SD 1.7; 101/294, 34.4%).

The results of the regression analyses with appreciation scores can be found in Table 5. The results showed that there were no demographic characteristics associated with appreciation scores. The only variable that was significantly associated with all appreciation scores was competence: feeling more confident because of the program in eating (more) healthily or engaging in sufficient PA was associated with higher appreciation scores. For the overall appreciation score, it was found that choosing both modules compared with no module was linked to a lower appreciation score. However, completing more sessions in the PA module was related to a higher appreciation score. For both the diet and PA appreciation scores, the results showed that feeling more related to the program was linked to higher appreciation scores. For the appreciation of the diet module, it was found that a higher amotivation to PA and being more intrinsically motivated to eat (more) healthily was linked to higher appreciation scores. For the appreciation of the PA module, it was found that a higher amotivation to eat (more) healthily was related to a lower appreciation score, whereas being more autonomously motivated to engage in sufficient PA was linked to a higher appreciation score. The full basic models can be found in Multimedia Appendix 2.
Table 5. Results of the stepwise regression analyses (full model) showing variables associated with appreciation scores.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Appreciation b</th>
<th>Appreciation of diet module c</th>
<th>Appreciation of PA d module e</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>y b (SE)</td>
<td>B b P value</td>
<td>b (SE) B P value</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.43 (1.17)</td>
<td>0.00 .71</td>
<td>1.17 (1.35) 0.00 .39</td>
</tr>
<tr>
<td>Age (years)</td>
<td>0.01 (0.01)</td>
<td>0.06 .35</td>
<td>-.00 (0.01) -.03 .71</td>
</tr>
<tr>
<td>Gender h</td>
<td>-.11 (0.18)</td>
<td>-.03 .56</td>
<td>-.15 (0.20) -.04 .45</td>
</tr>
<tr>
<td>Education high i</td>
<td>0.14 (0.21)</td>
<td>0.04 .50</td>
<td>0.05 (0.22) 0.01 .81</td>
</tr>
<tr>
<td>Education low i</td>
<td>0.28 (0.46)</td>
<td>0.03 .54</td>
<td>-.25 (0.43) -.03 .57</td>
</tr>
<tr>
<td>Marital partner k</td>
<td>0.21 (0.19)</td>
<td>0.06 .29</td>
<td>0.03 (0.21) 0.01 .88</td>
</tr>
<tr>
<td>Work k</td>
<td>0.0003 (0.19)</td>
<td>0.0001 .99</td>
<td>0.07 (0.20) 0.02 .71</td>
</tr>
<tr>
<td>Impairment l</td>
<td>-.73 (0.41)</td>
<td>-.10 .08</td>
<td>-.09 (0.39) -.01 .82</td>
</tr>
<tr>
<td>BMI</td>
<td>0.01 (0.02)</td>
<td>0.03 .55</td>
<td>-.01 (0.02) -.04 .59</td>
</tr>
<tr>
<td>Health status</td>
<td>0.006 (0.01)</td>
<td>0.04 .52</td>
<td>-.002 (0.01) -.02 .80</td>
</tr>
<tr>
<td>Amotivation diet</td>
<td>0.14 (0.11)</td>
<td>0.10 .20</td>
<td>-.10 (0.11) -.07 .38</td>
</tr>
<tr>
<td>Amotivation PA</td>
<td>0.0003 (0.11)</td>
<td>0.0002 .99</td>
<td>0.24 (0.12) 0.17 .05n</td>
</tr>
<tr>
<td>Controlled motivation Diet</td>
<td>0.07 (0.12)</td>
<td>0.05 .56</td>
<td>-.09 (0.15) -.07 .53</td>
</tr>
<tr>
<td>Controlled motivation PA</td>
<td>-.14 (0.12)</td>
<td>-.10 .27</td>
<td>-.09 (0.14) -.07 .51</td>
</tr>
<tr>
<td>Autonomous motivation diet</td>
<td>-.05 (0.14)</td>
<td>-.03 .74</td>
<td>0.25 (0.19) 0.17 .18</td>
</tr>
<tr>
<td>Autonomous motivation PA</td>
<td>0.23 (0.14)</td>
<td>0.15 .09</td>
<td>0.09 (0.19) 0.06 .64</td>
</tr>
<tr>
<td>Intrinsic motivation diet</td>
<td>0.19 (0.12)</td>
<td>0.10 .11</td>
<td>0.31 (0.13) 0.18 .02n</td>
</tr>
<tr>
<td>Intrinsic motivation PA</td>
<td>-.12 (0.10)</td>
<td>-.08 .23</td>
<td>-.09 (0.11) -.06 .40</td>
</tr>
<tr>
<td>Autonomy</td>
<td>0.22 (0.13)</td>
<td>0.10 .10</td>
<td>0.17 (0.15) 0.06 .27</td>
</tr>
<tr>
<td>Relatedness</td>
<td>0.31 (0.18)</td>
<td>0.16 .10</td>
<td>0.38 (0.17) 0.20 .03m</td>
</tr>
<tr>
<td>Competence</td>
<td>0.50 (0.14)</td>
<td>0.34 &lt;.001m</td>
<td>0.78 (0.13) 0.49 &lt;.001m</td>
</tr>
<tr>
<td>Diet advice green a</td>
<td>1.25 (1.43)</td>
<td>0.04 .38</td>
<td>__ .o ____ ____ ____</td>
</tr>
<tr>
<td>Diet advice red a</td>
<td>0.14 (0.21)</td>
<td>0.03 .49</td>
<td>-.07 (0.22) -.02 .74</td>
</tr>
<tr>
<td>PA advice green a</td>
<td>0.43 (0.61)</td>
<td>0.06 .48</td>
<td>0.04 (0.68) 0.01 .96</td>
</tr>
<tr>
<td>PA advice red a</td>
<td>-.32 (0.73)</td>
<td>-.04 .66</td>
<td>-.26 (0.81) -.03 .74</td>
</tr>
<tr>
<td>Module choice diet p</td>
<td>-.45 (0.31)</td>
<td>-.11 .14</td>
<td>-.26 (0.19) -.07 .18</td>
</tr>
<tr>
<td>Module choice PA p</td>
<td>0.32 (0.54)</td>
<td>0.03 .55</td>
<td>____ ____ 0.29 (0.70) 0.03 .68</td>
</tr>
<tr>
<td>Module choice both p</td>
<td>-.58 (0.27)</td>
<td>-.17 .03m</td>
<td>____ ____ ____ ____</td>
</tr>
<tr>
<td>Sessions diet</td>
<td>0.03 (0.07)</td>
<td>0.04 .61</td>
<td>-.04 (0.10) -.02 .67</td>
</tr>
<tr>
<td>Sessions PA</td>
<td>0.15 (0.07)</td>
<td>0.15 .04m</td>
<td>____ ____ 0.27 (0.16) 0.13 .09</td>
</tr>
</tbody>
</table>

aThe results’ interpretations are reported when all other predictors are held constant.
bObservations=291; \( R^2 / R^2\) adjusted=0.431/0.368; Akaike information criterion=1047.00.

cObservations=159; \( R^2 / R^2\) adjusted=0.669/0.607; Akaike information criterion=494.01.

dPA: physical activity.

eObservations=101; \( R^2 / R^2\) adjusted=0.726/0.630; Akaike information criterion=318.10.

f b: unstandardized regression coefficient.

g B: standardized regression coefficient.
**Discussion**

**Principal Findings**

This study has described the flow of participants in the MyLifestyleCoach intervention and identified characteristics related to this intervention’s use and appreciation. Our first aim was to describe the participants’ flow. Our findings resemble the typical nonusage attrition curve [3]. This was indicated by the largest drop in participation after the first session, and the attrition rate declined exponentially in subsequent sessions. More than half of the participants completed the entire 4 sessions in the case of 1 module or even 8 sessions in the case of both the diet and PA module intervention once they completed the first session of the module or modules. For instance, by implementing autonomy in our intervention and by offering participants a choice in which module or modules to participate, we expected to reduce this decline. Unfortunately, this was not the case, as only 20%-30% (diet module: 49/146, 33.6%; PA module: 2/12, 17%; both modules: 58/273, 21.2%) of participants completed their chosen module or modules. These numbers were (slightly) lower than those of other multisession PA interventions [34,35]. However, these interventions are not entirely comparable. Our intervention was more elaborate and did not only concern PA behavior but also concerned diet behavior. In addition, 2 findings regarding offering choices to the participants are worth mentioning; refer to the Implications section. When participants received a reminder email to revise their initial choice of following no module, it was remarkable that only 2 participants changed their initial module choice and decided to start a module. Another important finding was that more participants who indicated not to follow a module in the opening session were more likely to complete the follow-up questionnaire than participants who did not enter the opening session or did not make a module choice.

Our second aim was to examine which characteristics are linked to the use of initial and follow-up sessions. This study has demonstrated that age, marital status, health status, BMI, and physical impairment were related to key moments of stopping to use our intervention. In general, these findings are in line with previous literature [7,10,12,13,36]. Interestingly, a previous study by our research group discovered a trend that people with physical disabilities were less likely to choose the PA module on top of the diet module [22]. In this study, we found that they were more likely to complete the first session of the diet module when only the diet module was chosen. A reason for this could be that people with physical disabilities are more likely to focus and work on their dietary behavior, as they have fewer options to improve their PA because of their impairment. This finding shows the potential for specific groups in eHealth interventions in which people can choose the behavior or behaviors they prefer to work on. Furthermore, there seems to be a pattern that different motivation types are related to use at different points when people are likely to stop using the intervention. More autonomous motivation was associated with the completion of the opening session and the first session of the diet module. In contrast, participants who were more intrinsically motivated toward a healthy diet were less likely to complete these initial sessions. This finding is consistent with a study that found that users with higher levels of autonomous motivation toward dietary behaviors at baseline were less likely to stop using the intervention at an early moment [17]. Thus, perceiving eating healthily as a personally valued (and integrated) goal may be a relevant driver of initial use than the inherent joy of a healthy diet. On the other hand, more intrinsic motivation toward PA was associated with completing the opening session. This indicates that engaging in PA for inherent joy is relevant in initial use. Controlled motivation toward PA was linked to the starting and completion of both modules. These participants could have felt more pressure to start and complete the whole intervention by external regulations, for example, for a reward or introjected regulations for PA, for instance, to avoid negative feelings [37]. Thus, it is important to take the precise motivation type into account to stimulate use of the intervention.

Some program-related features are also linked to use, such as the advice a person received at the start of the intervention based on an assessment of the dietary and PA behavior on his or her initial behavioral performance. For instance, people with red advice for diet, indicating much room for improvement in their dietary behavior, were less likely to complete the whole intervention and thus both modules but were more likely to complete the diet module when only the diet module was chosen. The red diet advice seemed to increase participants’ focus to improve their PA because of their impairment. This finding shows the potential for specific groups in eHealth interventions in which people can choose the behavior or behaviors they prefer to work on. Furthermore, there seems to be a pattern that different motivation types are related to use at different points when people are likely to stop using the intervention. More autonomous motivation was associated with the completion of the opening session and the first session of the diet module. In contrast, participants who were more intrinsically motivated toward a healthy diet were less likely to complete these initial sessions. This finding is consistent with a study that found that users with higher levels of autonomous motivation toward dietary behaviors at baseline were less likely to stop using the intervention at an early moment [17]. Thus, perceiving eating healthily as a personally valued (and integrated) goal may be a relevant driver of initial use than the inherent joy of a healthy diet. On the other hand, more intrinsic motivation toward PA was associated with completing the opening session. This indicates that engaging in PA for inherent joy is relevant in initial use. Controlled motivation toward PA was linked to the starting and completion of both modules. These participants could have felt more pressure to start and complete the whole intervention by external regulations, for example, for a reward or introjected regulations for PA, for instance, to avoid negative feelings [37]. Thus, it is important to take the precise motivation type into account to stimulate use of the intervention.

Some program-related features are also linked to use, such as the advice a person received at the start of the intervention based on an assessment of the dietary and PA behavior on his or her initial behavioral performance. For instance, people with red advice for diet, indicating much room for improvement in their dietary behavior, were less likely to complete the whole intervention and thus both modules but were more likely to complete the diet module when only the diet module was chosen. The red diet advice seemed to increase participants’ focus to improve their PA because of their impairment. This finding shows the potential for specific groups in eHealth interventions in which people can choose the behavior or behaviors they prefer to work on. Furthermore, there seems to be a pattern that different motivation types are related to use at different points when people are likely to stop using the intervention. More autonomous motivation was associated with the completion of the opening session and the first session of the diet module. In contrast, participants who were more intrinsically motivated toward a healthy diet were less likely to complete these initial sessions. This finding is consistent with a study that found that users with higher levels of autonomous motivation toward dietary behaviors at baseline were less likely to stop using the intervention at an early moment [17]. Thus, perceiving eating healthily as a personally valued (and integrated) goal may be a relevant driver of initial use than the inherent joy of a healthy diet. On the other hand, more intrinsic motivation toward PA was associated with completing the opening session. This indicates that engaging in PA for inherent joy is relevant in initial use. Controlled motivation toward PA was linked to the starting and completion of both modules. These participants could have felt more pressure to start and complete the whole intervention by external regulations, for example, for a reward or introjected regulations for PA, for instance, to avoid negative feelings [37]. Thus, it is important to take the precise motivation type into account to stimulate use of the intervention.

The third aim of this study was to describe the appreciation scores for the MyLifestyleCoach intervention and examine its predictors. First, the participants rated the intervention as

---

\(^{h}\)Female is the reference category.

\(^{i}\)Medium education is the reference category.

\(^{j}\)Single is the reference category.

\(^{k}\)Being unemployed is the reference category.

\(^{l}\)No physical impairment is the reference category.

\(^{m}\)Values represent statistical significance.

\(^{n}\)Orange advice is the reference category.

\(^{o}\)These variables were not included in the model.

\(^{p}\)Choosing no module is the reference category in the general intervention’s appreciation, whereas choosing both modules is the reference category for the appreciation of the diet module and the physical activity module.
reasonably positive. Second, no demographic factors were related to the appreciation scores. Regarding the motivational constructs, being less motivated toward being physically active, thus having higher amotivation, was associated with a higher appreciation of the diet module. These participants might be solely interested in the diet module and give a higher rating as a result. Having lower amotivation toward eating healthily was associated with a higher appreciation of the PA module. Those participants might value a healthy diet or even a healthy lifestyle, and as a result, provide a higher rating. Furthermore, more autonomous motivation toward being physically active was associated with a higher appreciation of the PA module. In addition, higher scores on basic psychological needs, particularly competence and relatedness, were linked to more favorable ratings of the program. Finally, some program features were also related to higher appreciation scores, such as the choice option. We found that choosing no module compared with both modules in the opening session and finishing more sessions in the PA module were related to a higher appreciation score. These evaluation scores, including an evaluation of the basic psychological needs of autonomy, competence, and relatedness, were assessed at the 6-month follow-up questionnaire. Here, a large proportion of participants (426/775, 55%) dropped out (see the Limitations section).

Implications

Although our intervention structure may not necessarily be generalizable to other interventions, we provide some important implications that could be useful for the development of future interventions. First, we found that sending a generic reminder email to the participants whose initial choice was not to participate in the intervention minimally increased further intervention use, as only 2 participants revised their choice. Therefore, generic reminder emails are not recommended for this purpose. Instead, emails that contain new or different content or are tailored to specific characteristics, such as the extent of self-determination, might motivate people more to initiate a module [39,40].

Second, our results show that when people are given the option of beginning directly after the opening session or at a later moment, attrition rates are not negatively affected; however, this option does not improve use either. People possibly experience more autonomy by choosing the time point of using intervention parts, which might prevent them from stopping using the intervention and dropping out early. Thus, it can be assumed that providing participants with an option of when to start with the intervention is not detrimental for use.

Third, after people completed the first session of a module, about half of them finished the intervention comprising 4 or even 8 sessions spread over 3 months. It is recommended to make the first session of an intervention short and challenging and allow the person to choose small goals and achieve some success. However, more in-depth research is needed to examine why some individuals are more likely to adhere at particular moments within the intervention or give more favorable ratings. Further research should be undertaken to explore what could be improved to make participants more likely to adhere to the intervention. Examples of possible improvements could be more relevant content, better tailoring to specific groups, or using motivational interviewing to improve importance as early as possible.

Finally, the relative number of completed follow-up questionnaires was similar for the participants who chose to follow 1 or both modules and for those whose choice was to not follow any module. This latter group might have been more likely to stop using the intervention modules when this had been made obligatory and at risk of dropping out for the follow-up questionnaires. Thus, giving them a choice to start with the intervention, which is with a particular module, has prevented losing them to the follow-up questionnaires. This approach might have resulted in a slightly higher percentage of people who completed the follow-up questionnaire (349/775, 45% vs 409/987, 41.4%) compared with the previously tested single-behavior I Move intervention [34].

Limitations

There are several limitations worth mentioning. First, only self-reports were used to gather data. People could have responded in a more socially desirable way. For instance, they could have reported consuming more fruit and vegetables than they actually consume. This could have an effect on the received advice in the opening session [41]. Second, selection bias may have been present at some points of using attrition and dropout in this study. We cannot further investigate this as no information, such as demographics or motivation, is available for those who did not fill in the baseline questionnaire after randomization. It is likely that those who were not motivated to change their behavior more often dropped out. Third, this study focused on the theoretical framework of SDT, particularly focusing on motivation as a predictor of use and appreciation. Other psychosocial constructs, such as intention, could also be relevant to use. In a previous study, we found that these variables are highly correlated [22]. Therefore, we did not include these variables in our analyses to avoid multicollinearity. Finally, generalizability may be questioned, as a large part of our sample was highly educated. This is generally found in eHealth research (eg, the study by Rhodes et al [42]). Although our results demonstrate that education is not related to use at any point in this intervention, our predominantly highly educated sample could also have biased our findings. Nevertheless, future studies could aim to develop promotion strategies to attract more specific subgroups, such as less educated people with a less healthy lifestyle [43].

Conclusions

This process evaluation study adds meaningful information on the use and appreciation of a web-based, computer-tailored dietary and PA intervention—MyLifestyleCoach. The results indicate that different types of motivation that were examined in this study at play at other moments where people are likely to stop using the intervention, such as the initial session or sessions or completing the whole intervention. Appreciation was associated with several motivational constructs, such as amotivation and intrinsic motivation, and related to basic psychological needs, such as competence. We derived some practical implications for developing eHealth interventions that contain multiple health behaviors. For instance, we found that
about half of the participants ended the entire intervention once they finished the first session. Therefore, we recommend making the first session in a multiple-session intervention short, challenging, and rewarding and allow the person to choose small goals and achieve success.

Acknowledgments
This project was funded by an internal research fund of the Open University of the Netherlands.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Results from the stepwise logistic regression predicting completion of intervention components.

Multimedia Appendix 2
Results from the stepwise linear regression predicting the appreciation score for the whole intervention, the diet module, and physical activity module.

References


Abbreviations

MVPA: moderate to vigorous physical activity
PA: physical activity
RCT: randomized controlled trial
SDT: self-determination theory

©Juul M J Coumans, Anke Oenema, Catherine A W Bolman, Lilian Lechner. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.12.2021. 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.
Development of a Web-Based School Support System Within the AVATAR Project for Psychosocial Well-being in Adolescents: Pilot Feasibility Study

Francesca Mastorci, PhD; Paolo Piaggi, PhD; Gabriele Trivellini; Cristina Doveri; Anselmo Casu; Luca Bastiani, PhD; Irene Marinaro; Cristina Vassalle, PhD; Alessandro Pingitore, MD, PhD

1Clinical Physiology Institute, Consiglio Nazionale delle Ricerche Area della Ricerca di Pisa (CNR), Pisa, Italy
2Department of Information Engineering, University of Pisa, Pisa, Italy
3Fondazione Toscana Gabriele Monasterio, Pisa, Italy

Corresponding Author:
Alessandro Pingitore, MD, PhD
Clinical Physiology Institute
Consiglio Nazionale delle Ricerche Area della Ricerca di Pisa (CNR)
Via Moruzzi 1
Pisa, 56124
Italy
Phone: 39 050312605
Email: pingi@ifc.cnr.it

Abstract

Background: Health and well-being promotions are key points of educational programs for adolescents within schools. There are several health education programs mainly based on lifestyle habit changes; however, social and emotional dimensions should be considered within these educational strategies.

Objective: This study aimed to (1) develop a new web-based school support system to assess and analyze individual, classroom, and scholastic institute data on lifestyle habits, social context, emotional status, and scholastic performance; (2) create a web tool for managing the well-being of adolescents through a dynamic and personalized interface that provides immediate feedback that allows the school to monitor progress; and (3) evaluate, in a pilot study, the feasibility of this web-based school support system in order to build health programs that are specific to the needs of the studied population.

Methods: The AVATAR (a new purpose for the promotion and evaluation of health and well-being among healthy teenagers) method consists of integrating the information coming from different questionnaires. In particular, to allow planning didactic and educational actions based on the results obtained, the AVATAR approach allows subdivision of the results of the different observed variables and the 4 components into the following 3 percentile categories: modify, improve, and maintain. The AVATAR web platform was designed to collect data on lifestyle, emotional status, and social context from junior high schools in terms of the fundamental aspects of adolescent daily life, with free use by the scholastic community (scholars, teachers, and parents). In this pilot/feasibility study, data from 331 students were acquired between 2018 and 2019 at the beginning of the scholastic year (pre) and at the end following the school-based program (post).

Results: Preliminary results showed that after school planning and specific program implementation, defined after AVATAR feedback, students reported better well-being perception characterized by higher perception in psychological well-being ($P=.001$), mood ($P=.001$), self-perception ($P=.006$), and autonomy ($P=.001$), and an increase in the perception of financial resources ($P=.001$), which helped in developing healthy lifestyle habits ($P=.007$). In the social context assessment, students reported stronger relationships with family ($P=.02$) and peers ($P=.001$), and a lower perception of bullying ($P=.001$).

Conclusions: The AVATAR web-based platform is a feasible and flexible tool for the health and well-being management of adolescents from epidemiological, preventive, and educational points of view. In particular, it can be used to (1) promote information campaigns aimed at modifying risk behaviors in the student population, (2) sensitize students and put them at the center of their growth path, (3) inform institutions about the health and well-being of the school population, (4) ensure health programs are acceptable and feasible to users before launching on a large scale, and (5) improve the relationship of users (school) and educational agencies with research groups.
adolescent; well-being management; schools; web tool; health promotion

Introduction

Health and well-being promotions are key points of educational programs for adolescents within schools. Schools allow obtaining a large number of adolescents from different family social backgrounds and represent a more naturalistic and interactive environment to induce positive effects on health from social (eg, family, peers, and teachers) and cultural perspectives [1,2]. There are several health education programs, which are mainly based on lifestyle changes (in particular, diet and exercise), that significantly reduce the incidences of obesity and metabolic syndrome [3]. However, considering that different factors have impacts on health and well-being in adolescents, including those in the social, emotional, and mental context, these health components should be implemented within health education programs [4]. In this regard, the KIDSCREEN questionnaire is a standardized tool to assess the quality of life and well-being of adolescents, which has been validated in several European countries and includes items of different areas impacting health in adolescents, that can help to identify critical areas for health education program interventions [5]. In this way, health education programs can be more focused on the needs of adolescents through the identification of the items to improve and those to potentiate. In the perspective of an integrated and multidisciplinary framework of health intervention, our group developed an integrated and personalized index of well-being built on the integration of the weights of items belonging to the dimensions of lifestyle, social context, emotional status, and mental skills [6]. This index has been built for a single adolescent, where health interventions can be oriented in a personalized approach. However, in a school context, health programs need to be oriented to the class and institute community, taking into account the health profile of each student. This pushed us to develop a tool in which the data of students on the abovementioned dimensions were provided as classroom data and institution data.

This study aimed to examine the potential impact of the development of a web-based school support system to acquire, archive, and analyze online data about lifestyle habits, social context, emotional status, and scholastic performance in order to provide immediate feedback of the results to the school. The goal of this pilot study was to evaluate the feasibility of this web-based tool to monitor longitudinally the effects of health programs adopted in relation to questionnaire results.

Methods

Overview

In this pilot/feasibility study, first, we developed a web-based school support system to acquire, archive, and analyze in real time data about lifestyle habits, social context, emotional status, and scholastic performance among students. Second, we reported the methods and results of the feasibility study. All procedures performed in this study were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The protocol was approved by the regional ethics committee (166/2018). In addition, the pilot/feasibility study was approved by the internal ethics committee of each participating school, in accordance with Italian law. All parents or legal guardians gave informed consent and authorized researchers to use the data in accordance with Italian law.

General Aspects of the AVATAR Web-Based Tool

The web-based platform was designed to collect data from junior high schools participating in the AVATAR (a new purpose for the promotion and evaluation of health and well-being among healthy teenagers) project on the fundamental aspects of adolescent daily life, with free use without a commercial license by the scholastic community (scholars, teachers, and parents). In particular, variables on health and well-being that encompassed different dimensions of health, including lifestyle, emotional status, and social context, were monitored. The platform contained questionnaires; training documents; and reports accessible to schools, teachers, parents, and people active in the field of education and prevention. This allowed us to (1) define and understand the needs of the population (adolescence) for interventions or public health programs, (2) evaluate if the proposed programs are acceptable and feasible for users before launch through a pilot study, and (3) improve the relationship between users (at school) and research groups.

Data Collection

Data from students at different times during the scholastic year (usually at the beginning and at the end of the scholastic year) were collected using the AVATAR web tool [7]. A sociodemographic data record was used to acquire information about gender, age, schooling, family structure, and BMI, according to World Health Organization age groups [8]. The Italian version of KIDSCREEN-52 was used to assess health-related quality of life [9,10]. The KIDSCREEN is a self-report questionnaire designed to assess health-related quality of life, with the aim to monitor and measure personal experiences in children and adolescents about their perceptions of health status and well-being. The questionnaire, which describes physical, psychological, mental, social, and functional aspects of well-being, consists of 52 items grouped into 10 dimensions [9,10]. The KIDSCREEN questionnaire has been psychometrically tested using data obtained in a multicenter European study that included a sample of 22,827 children recruited in 13 countries [11]. Dietary habits were evaluated using the Mediterranean Diet Quality Index for children and adolescents (KIDMED) [12]. The KIDMED index is based on principles sustaining Mediterranean dietary patterns, as well as those that undermine it. The index ranges from 0 to 12, and consists of a self-administered 16-question test. Physical activity
levels were assessed using the Physical Activity Questionnaire for Older Children (PAQ-C). The questionnaire provides a general measure of physical activity for those aged 8 to 20 years. The PAQ-C is a self-administered questionnaire consisting of 9 items rated on a 5-point scale. A higher score indicates more active children/adolescents [13].

The perception of school engagement was estimated through questions concerning scholastic achievements in language and literature, language acquisition, and science.

**Health-Related Quality of Life Components**

In the AVATAR platform, the following 4 components of health-related well-being have been considered: lifestyle habits, emotional status, social context, and mental skills.

<table>
<thead>
<tr>
<th>Component</th>
<th>Well-being dimensions (observed variables)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical well-being</td>
<td>Autonomy</td>
</tr>
<tr>
<td></td>
<td>Financial resources</td>
</tr>
<tr>
<td></td>
<td>Diet</td>
</tr>
<tr>
<td></td>
<td>Physical activity</td>
</tr>
<tr>
<td>Parent relations</td>
<td></td>
</tr>
<tr>
<td>Peers</td>
<td></td>
</tr>
<tr>
<td>School environment</td>
<td></td>
</tr>
<tr>
<td>Bullying</td>
<td></td>
</tr>
<tr>
<td>Psychological well-being</td>
<td></td>
</tr>
<tr>
<td>Mood</td>
<td></td>
</tr>
<tr>
<td>Self-perception</td>
<td></td>
</tr>
<tr>
<td>Emotion</td>
<td></td>
</tr>
</tbody>
</table>

**Data Management**

The AVATAR method involves integrating the information coming from the different questionnaires. In particular, to allow planning didactic and educational actions based on the results obtained, the AVATAR approach allows subdivision of the results of the different observed variables and the 4 components into the following 3 percentile categories: modify, improve, and maintain.

To obtain this representation of the data, for each investigated area of the questionnaire, the cutoffs for adolescent-reported dimension and total scores were defined using the 10th and 90th percentiles, based on the sample distribution into 3 categories. Scoring in the ≤10th percentile was used to classify the proportion of adolescents having poor quality in each investigated area (orange; modify), scoring in the ≥90th percentile was used to classify the proportion of adolescents having high quality (blue; improve), and scoring between the 10th and 90th percentiles was used for intermediate values (green; maintain) [19,20].

The teacher can view the data expressed as a single variable or merged and integrated into the 4 components of the entire class and the institution (Figure 1), up to the individual pupil, identified with an ID (Figure 2).
Methods of the Pilot/Feasibility Study

In order to evaluate the feasibility of the AVATAR web-based school support system, a pilot study, as a part of the AVATAR project, was performed. Data collection was conducted between 2018 and 2019 from 1 of 10 junior high schools at the beginning (pre) and at the end (post) of the school year. In total, 331 students (172 female students, 52%; 159 male students, 48%;
mean age 12.5 years, SD 1 year) were included. Adolescent students were enrolled according to the following inclusion criteria: age 10 to 14 years, absence of neuropsychiatric or other diseases, informed consent signed, and completion of the questionnaires proposed. In every school class, all adolescents filled out the questionnaires, and those who were not eligible were excluded from the study retrospectively. The questionnaires, which have been previously described, were filled at the beginning and end of the school year.

Participants were previously instructed on how to fill out the questionnaires and how to conduct the tests. All tests were conducted during participants’ computer lessons in school. No incentive was provided to adolescents or parents. A research assistant was available to provide information and technical support to complete the questionnaires.

School-Based Program Built on the AVATAR Platform

With the results obtained by the AVATAR platform in the first AVATAR administration (at the beginning of the school year), school planning and specific program implementation were aimed to support the development of students’ emotional-relational skills and competences. In particular, the project and didactic actions were oriented on the basis of the identified needs and aimed to (1) strengthen the self-efficacy of students, as well as their social skills, to increase personal responsibility in relationships for developing autonomy (empathy, self-regulation, and self-efficacy) and (2) enhance citizenship skills (theatre, environment, art, counselling, nutrition, sports, breathing, relaxation, etc) with a view of prevention, which was integrated with curricular planning. For these reasons and for didactic/educational aims, the pilot study addressed the following 3 AVATAR areas: lifestyle habits, social context, and emotional status.

Statistical Analysis

Statistical data analyses were performed using SPSS (version 22.0; IBM Corp). Data are presented as mean (SD) or as mean with 95% CI. Alpha was set at .05, and 2-sided P values have been reported. Changes in health-related quality of life (baseline vs post) were analyzed using the Student paired t test. The McNemar-Bowker test was used to evaluate changes in the proportion of subjects belonging to each tertile of health-related quality of life variables after the intervention.

Results

Effects of the School-Based Program on Health-Related Quality of Life and Lifestyle Habits

Descriptive data on health-related quality of life and lifestyle habits (diet and physical activity; from preprogram and postprogram) are presented in Table 2. Data on the KIDSCREEN-52 dimensions are calculated as mean T-scores according to the KIDSCREEN Group [9,10]. During the school year, following the program built on the results of the first administration, students showed a higher perception in psychological well-being (P=.001), mood (P=.001), self-perception (P=.006), and autonomy (P=.001), which was understood as the opportunity to create his/her social and leisure time. There was an increase in the well-being perception owing to the school-based program, as well as the perception of financial resources compared with the initial situation (P=.001).

In the social context, students reported higher values in the relationship with their family (P=.02) and peers (P=.001), and exhibited a lower perception of bullying (P=.001). For lifestyle habits, after the personalized program, students developed higher adherence to the Mediterranean diet (P=.007) and higher physical activity levels (P=.001) compared with the previous condition.
Table 2. Questionnaire findings in the pilot study sample at the beginning (pre) and end (post) of the school year.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Pre score (n=331), mean (SD)</th>
<th>Post score (n=331), mean (SD)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>KIDSCREEN-52&lt;sup&gt;b&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical well-being</td>
<td>47.35 (SD 18.52)</td>
<td>47.61 (SD 22.05)</td>
<td>.85</td>
</tr>
<tr>
<td>Psychological well-being</td>
<td>44.50 (SD 15.86)</td>
<td>52.59 (SD 9.67)</td>
<td>.001</td>
</tr>
<tr>
<td>Mood/emotion</td>
<td>46.95 (SD 13.52)</td>
<td>50.71 (SD 10.76)</td>
<td>.001</td>
</tr>
<tr>
<td>Self-perception</td>
<td>51.45 (SD 15.8)</td>
<td>53.79 (SD 11.32)</td>
<td>.006</td>
</tr>
<tr>
<td>Autonomy</td>
<td>45.21 (SD 14.21)</td>
<td>50.96 (SD 10.85)</td>
<td>.001</td>
</tr>
<tr>
<td>Parent relationship</td>
<td>51.55 (SD 11.52)</td>
<td>53.08 (SD 10.27)</td>
<td>.02</td>
</tr>
<tr>
<td>Financial resources</td>
<td>47.66 (SD 13.44)</td>
<td>52.20 (SD 8.95)</td>
<td>.001</td>
</tr>
<tr>
<td>Peers</td>
<td>50.15 (SD 12.04)</td>
<td>53.06 (SD 11.01)</td>
<td>.001</td>
</tr>
<tr>
<td>School environment</td>
<td>50.34 (SD 10.43)</td>
<td>51.24 (SD 9.38)</td>
<td>.07</td>
</tr>
<tr>
<td>Social acceptance</td>
<td>46.02 (SD 15.22)</td>
<td>49.57 (SD 10.67)</td>
<td>.001</td>
</tr>
<tr>
<td><strong>KIDMED&lt;sup&gt;c&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.76 (SD 2.42)</td>
<td>6.13 (SD 2.45)</td>
<td></td>
<td>.007</td>
</tr>
<tr>
<td><strong>PAQ-C&lt;sup&gt;d&lt;/sup&gt;</strong></td>
<td>2.57 (SD 0.67)</td>
<td>2.76 (SD 0.66)</td>
<td>.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>P values were calculated using the Student paired t test.

<sup>b</sup>Data on the KIDSCREEN-52 dimensions were calculated as mean T-scores according to the KIDSCREEN Group.

<sup>c</sup>KIDMED: Mediterranean Diet Quality Index for children and adolescents.

<sup>d</sup>PAQ-C: Physical Activity Questionnaire for Older Children.

Effect of the School-Based Program on AVATAR Modify, Improve, and Maintain Percentile Categories

Descriptive data of health-related quality of life and lifestyle habits in the pre and post conditions expressed in percentage with respect to the modify, improve, and maintain percentile categories are presented in Table 3. After school planning and specific program implementation, the physical well-being dimension in the post condition changed, with an increase in the maintain percentile category and a reduction in the modify and improve percentile categories (Δ: P=.001). Psychological well-being perception was augmented in the maintain percentile category, and showed a decrease in the modify percentile category and an enhancement in the improve percentile category (Δ: P=.001). The emotion/mood dimension also changed after the school-based program, and this was characterized by a drop in the modify and improve percentile categories in the face of an increase in the maintain percentile category (Δ: P=.03). Autonomy, understood as the opportunity to create his/her social and leisure time, improved owing to the specific program, with an increase in the maintain percentile category and a significant decrease in the other percentile categories (Δ: P=.001). Moreover, perception of financial resources showed a decrease in the modify percentile category and an increase in the improve percentile category, and demonstrated stability in the maintain percentile category (Δ: P=.001). In the social context component, peer relationships showed an increase in the maintain percentile category and a decrease in the modify and improve percentile categories (Δ: P=.003). Lastly, the school-based program created according to the precondition results resulted in an improvement in the perception of social acceptance. In fact, for the bullying dimension, there was a significant reduction in the modify percentile category and an increase in the improve percentile category (Δ: P=.01).
Table 3. Changes in questionnaire findings (pre vs post) according to the modify, improve, and maintain percentile categories.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Modify percentile category, %</th>
<th>Improve percentile category, %</th>
<th>Maintain percentile category, %</th>
<th>Change (Δ), %</th>
<th>P valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre Post</td>
<td>Pre Post</td>
<td>Pre Post</td>
<td>Pre Post</td>
<td>Post Post</td>
</tr>
<tr>
<td>KIDSCREEN-52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical well-being</td>
<td>18.4</td>
<td>18.1</td>
<td>67.4</td>
<td>50.0</td>
<td>14.2</td>
</tr>
<tr>
<td>Psychological well-being</td>
<td>27.5</td>
<td>7.6</td>
<td>66.8</td>
<td>78.5</td>
<td>5.7</td>
</tr>
<tr>
<td>Mood/emotion</td>
<td>8.8</td>
<td>5.4</td>
<td>78.5</td>
<td>76.1</td>
<td>12.7</td>
</tr>
<tr>
<td>Self-perception</td>
<td>12.1</td>
<td>10.3</td>
<td>87.9</td>
<td>89.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Autonomy</td>
<td>10.6</td>
<td>4.5</td>
<td>77.3</td>
<td>72.5</td>
<td>12.1</td>
</tr>
<tr>
<td>Parent relationship</td>
<td>8.8</td>
<td>8.2</td>
<td>91.2</td>
<td>91.8</td>
<td>0.0</td>
</tr>
<tr>
<td>Financial resources</td>
<td>13.3</td>
<td>4.5</td>
<td>86.7</td>
<td>95.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Peers</td>
<td>10.0</td>
<td>6.3</td>
<td>82.2</td>
<td>78.2</td>
<td>7.9</td>
</tr>
<tr>
<td>School environment</td>
<td>10.3</td>
<td>6.9</td>
<td>77.9</td>
<td>78.2</td>
<td>11.8</td>
</tr>
<tr>
<td>Social acceptance</td>
<td>13.3</td>
<td>7.9</td>
<td>86.7</td>
<td>92.1</td>
<td>0.0</td>
</tr>
<tr>
<td>KIDMEDb</td>
<td>9.4</td>
<td>7.3</td>
<td>76.7</td>
<td>78.2</td>
<td>13.9</td>
</tr>
<tr>
<td>PAQ-Cc</td>
<td>17.8</td>
<td>12.4</td>
<td>80.4</td>
<td>86.4</td>
<td>1.8</td>
</tr>
</tbody>
</table>

aP value is for the change (Δ).
bKIDMED: Mediterranean Diet Quality Index for children and adolescents.
cPAQ-C: Physical Activity Questionnaire for Older Children.

Discussion

Principal Findings

We explored the development of a web-based school support system for the promotion of health and well-being in adolescents, through a pilot/feasibility study conducted in a small sample of subjects before launching a larger study, thus allowing the definition of public health programs based on the specific needs of the studied population and improving the relationship between schools and research groups.

The AVATAR platform, which processes data automatically and returns the findings legibly and in real time, offers the possibility for schools to have updated and comparable data on adolescents’ behaviors and well-being perceptions to create a network of collaborations with multidisciplinary experts for the development of prevention and health promotion interventions focused on the actual needs detected, which could help in selecting best practices and organizing targeted training sessions.

In addition, according to the technological perspective, the AVATAR platform, which returns the data to the school classified as “modify, improve, and maintain,” is highly flexible and adaptable in its potential applications. This methodological innovation, linked to the innovative approach, involves the simplicity and usability of the data acquired [7]. Teachers, in fact, can directly visualize the classroom and the cluster of students with similar profiles, and analyze a personalized well-being index that provides an integrated and personalized perspective of adolescents’ well-being [6]. As shown by the pilot study, this model helps teachers to more appropriately select interventions and educational programs for individual students and for the classroom, and thus, can monitor compliance and effectiveness. In particular, the data available to the school, in addition to enhancing the school’s success, have preventive purposes and can help improve resilience, happiness, social involvement, self-esteem, and sociability [14,15]. Furthermore, the results obtained by the AVATAR platform allow teachers to measure objectively the perception of students’ health and well-being, and follow their evolution longitudinally in order to reduce risk behaviors or potential risk factors, in line with the primordial prevention statement, which represents one of the aims of the AVATAR platform. The same notion is suitable for promoting social and emotional health in the school context. In fact, the school engaged in the pilot study, in agreement with school health educators who have increased skill development in the dimension of decision-making, had planned specific programs designed to support these areas of learning as closely related to school success and to strengthen relationships with family and peers [21]. This choice, according to the needs that emerged from the first monitoring, is linked to the notion that if adolescents are not conscious of their feelings and emotions, they will find it difficult to make reasoned choices, to choose healthy behaviors, and to achieve a good degree of learning [22].

Health-related quality of life and lifestyle habit data obtained after the school-based program showed better perceptions in the psychological and physical well-being, mood/emotion, self-perception, autonomy, and financial resources dimensions. In the social context, relationships with family and peers, as well as the perception of bullying, improved after didactic actions oriented toward these dimensions. Moreover, in the lifestyle assessment, owing to the school project, adolescents...
developed higher adherence to a Mediterranean diet and a better physical activity level.

When the data are analyzed according to the AVATAR approach, which involves partitioning into percentiles, what emerges is an improvement in physical and psychological well-being perceptions, emotional responses, and social acceptance with peers, with an increased percentage in the maintain percentile category and a reduction in the modify and improve percentile categories. These results obtained in the pilot study are in line with evidence that considers these variables responsible for healthy behavior and better health-related quality of life [23].

However, despite the objective well-being brought by a specific program aimed to support and potentiate the development of students’ emotional-relational skills and competences, when schools adopt cross-curricular programs, they face different implementation difficulties, such as lack of systematcity and objectivity.

Nevertheless, there are indications, in fact, that school-based programs based on educational activities augment self-image and body satisfaction in adolescents aged 12 to 14 years, probably because the school represents a good platform for increasing empowerment and awareness to health and well-being perceptions [24,25]. According to an ecological model, school-based programs, oriented to increase health status, should consider that school performance not only is the result of learning, but also, above all, depends on the environment in which the student lives and relates, and therefore, on the family microsystem and family-school macrosystem. In this new scenario, the adolescent must become aware of his/her own well-being and have at the same time (by the school) the tools to be proactive and participatory.

To modify individual unhealthy behaviors, it is necessary, therefore, to create environmental conditions suitable for encouraging a healthy lifestyle through an “intersectoral” and transversal approach to risk factors, considering all levels of intervention, both social and psychological. For this, there is a need for more evidence-based school intervention programs and an accurate assessment of their overall efficiency and efficacy.

**Strengths and Limitations**

A key strength of the AVATAR platform is that it is the first platform, to our knowledge, designed to introduce a web-based health promotion tool for the school community [26]. Importantly, monitoring the results allows the identification of the strength and fragile characteristics of each adolescent in order to define personalized educational programs. In particular, with the data, the school can (1) carry out longitudinal monitoring of students, (2) orient didactic actions, and (3) promote the personalization of educational and training courses.

In this context, the AVATAR platform intervenes at multiple points in the educational process in young people by combining management and empowerment of health and well-being, applying prevention strategies to reduce disease burden and health expenditure in adulthood, and enhancing learning. The AVATAR platform combines the rigor and objectivity of scientific and technological research with the needs that emerge from the school system regarding health, well-being, and educational success, integrating primordial prevention with the definition of a model of support for the autonomous design of schools. The development of a school-based program from the AVATAR web-based platform represents a powerful pattern for the promotion of health and psychosocial well-being, in which a network of different stakeholders dedicated to education may cooperate together to increase awareness, reduce risk behaviors, and potentiate educational success. Moreover, the AVATAR web-based tool offers a personalized well-being index that may allow the adoption of more individually focused strategies and interventions to improve well-being. Finally, as the AVATAR platform is delivered to students via web-based technology with interactive components through the involvement of teachers, student engagement and program fidelity can be increased, and this is supported by new national and European policies on the welfare of adolescents.

Several limitations should be acknowledged. First, since the questionnaires were completed during a school class, the environment may have biased the students’ responses. Finally, a control group was not included to demonstrate the effectiveness of the targeted intervention based on the needs identified because all schools, due to institutional obligations, carry out projects that could, in any case, impact the sampling results at the end of the school year.

**Conclusion**

The AVATAR monitoring platform, as shown by our pilot results, is configured as a tool for enhancing school autonomy, and offers schools and the community the possibility of having updated and comparable data on adolescents’ behaviors and perceptions to create a network of collaborations with multidisciplinary experts for the development of prevention and health promotion interventions focused on the actual needs identified, which could help in selecting best practices and organizing targeted training sessions. Owing to its flexibility and adaptability, the AVATAR platform can be used to (1) promote information campaigns aimed at modifying risk behaviors in the student population, (2) sensitize students and put them at the center of their growth path, (3) inform institutions on the health and well-being of the school population, (4) evaluate the effectiveness of the actions implemented by schools, and (5) promote the exchange of good practices aimed at strengthening systems intended for the education, training, and well-being of students.

**Acknowledgments**

We thank the Italian schools that joined the Italian school network “RETE ULISSE. Scuole insieme per la ricerca scientifica e l’innovazione didattica.” We particularly thank Istituto Comprensivo P Aldi (Grosseto) for pilot study monitoring.
Conflicts of Interest
None declared.

References


Abbreviations

AVATAR: a new purpose for the promotion and evaluation of health and well-being among healthy teenagers
KIDMED: Mediterranean Diet Quality Index for children and adolescents
PAQ-C: Physical Activity Questionnaire for Older Children

©Francesca Mastorci, Paolo Piaggi, Gabriele Trivellini, Cristina Doveri, Anselmo Casu, Luca Bastiani, Irene Marinaro, Cristina Vassalle, Alessandro Pingitore. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.12.2021. 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.
Development of the Shift Smartphone App to Support the Emotional Well-Being of Junior Physicians: Design of a Prototype and Results of Usability and Acceptability Testing

Isabelle Counson¹,², MPH, GradDipPsychAdv; Alexandra Bartholomew¹, BCom (BSc), MPH; Joanna Crawford¹,², PhD; Katherine Petrie¹,², BSc (Hons); Geetanjali Basarkod³, PhD, BSc (Hons); Victoria Moynihan¹, BVA; Josie Pires¹, BSc; Rachel Cohen¹, PhD, BSc (Hons), MPsych; Nicholas Glozier⁴, MBBS, FRANZCP, PhD; Samuel Harvey¹,², MBBS, MRCGP, MRCPSych, FRANZCP, PhD; Samineh Sanatkar¹,², PhD

¹Black Dog Institute, Randwick, Australia
²School of Psychiatry, Faculty of Medicine, UNSW Sydney, Kensington, Australia
³Institute for Positive Psychology and Education, Australian Catholic University, North Sydney, Australia
⁴Brain and Mind Centre, Faculty of Medicine and Health, University of Sydney, Australia

Corresponding Author:
Samineh Sanatkar, PhD
Black Dog Institute
Hospital Road
Randwick, 2031
Australia
Phone: 1 02 9382 4368
Email: s-sanatkar@unsw.edu.au

Abstract

Background: Junior physicians report higher levels of psychological distress than senior doctors and report several barriers to seeking professional mental health support, including concerns about confidentiality and career progression. Mobile health (mHealth) apps may be utilized to help overcome these barriers to assist the emotional well-being of this population and encourage help-seeking.

Objective: This study describes the development and pilot trial of the Shift mHealth app to provide an unobtrusive avenue for junior physicians to seek information about, and help for, well-being and mental health concerns, which is sensitive to workplace settings.

Methods: A 4-phase iterative development process was undertaken to create the content and features of Shift involving junior physicians using the principles of user-centered design. These 4 phases were—needs assessment, on the basis of interviews with 12 junior physicians; prototype development with user experience feedback from 2 junior physicians; evaluation, consisting of a pilot trial with 22 junior physicians to assess the usability and acceptability of the initial prototype; and redesign, including user experience workshops with 51 junior physicians.

Results: Qualitative results informed the content and design of Shift to ensure that the app was tailored to junior physicians’ needs. The Shift app prototype contained cognitive behavioral, mindfulness, value-based actions, and psychoeducational modules, as well as a tracking function that visualized patterns of daily variations in mood and health behaviors. Pilot-testing revealed possible issues with the organization of the app content, which were addressed through a thorough restructuring and redesign of Shift with the help of junior physicians across 3 user experience workshops.

Conclusions: This study demonstrates the importance of ongoing end user involvement in the creation of a specialized mHealth app for a unique working population experiencing profession-specific stressors and barriers to help-seeking. The development and pilot trial of this novel Shift mHealth app are the first steps in addressing the mental health and support-seeking needs of junior physicians, although further research is required to validate its effectiveness and appropriateness on a larger scale.

(JMIR Form Res 2021;5(12):e26370) doi:10.2196/26370
KEYWORDS
digital mental health; mHealth apps; help-seeking; junior physicians; co-design; user-centered design; mobile phone

Introduction

Junior physicians exhibit levels of psychological distress and emotional exhaustion to a greater degree than their senior counterparts [1,2]. Junior physicians reported feeling impacted by a range of workplace-related stressors, including long working hours, a lack of breaks, and, at times, bullying and harassment [3,4]. A review of several prospective studies showed that individual factors, such as self-criticism, emotional instability, and a family history of psychopathology, are predictive of mental illness in junior physicians [5]. Furthermore, structural and personal barriers to help-seeking for mental health concerns have been noted, such as concerns about being reported to medical regulators and anxiety about reversing roles from being a physician to becoming a patient [6-8]. Australian data indicate that the most commonly reported barriers to professional help-seeking for depression among depressed physicians are privacy and confidentiality concerns [9].

Delays in receiving targeted treatment have the potential to compound and prolong symptoms of poor mental health as well as to increase the likelihood of developing comorbidities such as alcohol dependence [7,10]. Research suggesting a negative association between physicians’ psychopathology and best patient care practices highlights the implications of physicians’ mental health in the broader community [11-13]. Therefore, in addition to the adverse effects on the individual, it is in the wider public interest to support junior physicians in their transition into a demanding work environment and to help them seek and receive effective mental health care.

Although psychological interventions designed specifically to support physicians’ mental health are scarce, a recent meta-analysis found that interventions targeting physicians yielded small but significant reductions in symptoms of common mental disorder and suicidal ideation, particularly when therapeutic components were based on principles of cognitive behavioral therapy and mindfulness [14,15]. Of note, one study found that medical interns randomly assigned to a web-based cognitive behavioral therapy intervention group were 60% less likely to report suicidal ideation during their internship year than the comparator attention-control group [16]. Further research suggests that valued living and present-moment awareness components practiced in acceptance and commitment therapy and psychoeducational programs may be useful in reducing psychological distress in health care professionals and students [17-19].

Although the potential benefits of teaching these skills to junior physicians are clear, the practicality of such interventions can be challenging. Junior physicians are busy and regularly move between roles. New digital tools may be able to assist with these logistical challenges. Although mobile health (mHealth) apps have been shown to improve working populations’ mental well-being and willingness to seek help for mental health concerns [20-23], they have yet to be tested in well-controlled studies among physicians. Many physicians have already used mobile apps in the workplace to guide their management or prescription. A novel mHealth app using the same avenue as these professional development tools may be an acceptable way of delivering mental health support to junior physicians.

Previous research highlights the importance of a user-centered approach to developing mHealth app interventions [24-26] and, if designed for employees, additional factors, such as the workplace environment, should be taken into account [27,28]. As such, mHealth apps involving workplace considerations pose additional constraints on the app development process to ensure that content is effective, adequately delivered, and suits the target working population. User-centered design approaches involve iterative phases of prototyping, ideally employing co-design and end user feedback, and involve consideration of the user at every stage of the design process to ensure that the intervention meets their needs [29,30]. This, in turn, has been reported to maximize users’ engagement with, and adherence to, an mHealth intervention, and hence its impact [31].

This paper describes the process of developing Shift, a self-guided mental health and help-seeking smartphone app for junior physicians located in New South Wales, Australia. To our knowledge, this is the first mHealth app designed specifically to support the mental health of junior physicians. We used a user-driven and iterative development process, employing the principles of user-centered design and a multiphase process. This paper presents the 4-phase Shift app development process and how this process incorporates knowledge and feedback derived from qualitative assessments, pilot-testing, stakeholder and expert consultations, and user experience workshops with the target population.

Methods

Overview

There were 4 project phases as follows: phase 1, needs assessment through qualitative end user interviews; phase 2, Shift app prototype development; phase 3, pilot-testing of the Shift app prototype; and phase 4, generation of an updated version of the Shift app. In preparation for phase 1, consultations with a range of stakeholders were conducted (including junior physician managers, providers of support services for junior physicians, and professional organizations related to junior physicians in New South Wales) to examine the existing mental health support services for junior physicians, and facilitators of and barriers to engagement with these services. This was to ensure that the app development complemented existing support and provided up-to-date information on available services. Furthermore, our broader research team had previously developed HeadGear, an mHealth app that has been found to be effective in male-dominated working populations [21,32,33]. The intervention component in HeadGear was delivered in a 30-day challenge format, which successively unlocked psychoeducational material, as well as behavioral activation, goal-setting, and mindfulness techniques. Hence, one option
available to our team was to utilize the 30-day challenge format of the evidence-based HeadGear app and to modify the clinical content to meet the needs of junior physicians. To this end, the interviews with junior physicians in phase 1 included a question about their attitudes toward a 30-day challenge in an app to support the well-being of junior physicians.

**Phase 1: Needs Assessment Through Qualitative Interviews**

The objective of the qualitative component of this project was to inform the design and development of clinical content and to ensure that the app is tailored to the specific needs, characteristics, and challenges faced by this unique user group. Specifically, the interviews aimed to (1) identify the main stressors and challenges for junior physicians, both at and outside of work; (2) explore their attitudes toward a mental health app; (3) identify facilitators of and barriers to their use of, interest, and engagement in a mental health app; (4) assess current use of general apps; and, finally, (5) to identify their suggestions, preferences, and dislikes or unwanted features that they felt would support junior physicians’ mental health and well-being. The qualitative component was granted full ethical approval by the South Eastern Sydney Local Health District Human Research Ethics Committee (protocol #: 18/140, HREC/18/POWH/321). A qualitative analysis of the interviews focusing on the experiences of junior physicians is reported in detail elsewhere [34]. This paper presents specific app-related items and findings from interviews.

Participants were recruited from 2 metropolitan hospital sites in the Sydney area between July and September 2018. Advertising took place via email invitations distributed by medical supervisors, through email and social media announcements to professional organizations related to junior physicians, and through on-site hospital visits by members of the research team (JC, KP, and GB); 41 junior physicians expressed interest in the study, of whom 12 provided written informed consent and were recruited for one-on-one face-to-face interviews (conducted by JC). Participants included 9 women and 3 men aged between 24 and 35 years. The sample was evenly spread across the early stages of training and comprised interns (n=3), residents (n=5), and registrars (n=4). The majority had studied medicine in Australia and were generally 3-5 years into their clinical training.

Data on app-related items were analyzed using a thematic analysis approach [35] informed by grounded theory and Massey emergent approach [36]. In an iterative process undertaken by one researcher (KP), all transcripts were reviewed closely to generate an initial first-level coding framework. Through subsequent refinements through discussion with co-authors, broader second-level themes related to the app issues were identified, and subthemes and common suggestions were listed under each of the 4 main issues of interest, summarized below in the Results section.

**Phase 2: Prototype Development**

On the basis of the recommendations made by junior physicians in the qualitative interviews and consultations with stakeholders in line with evidence from recent literature [14,17,18], new clinical content for the Shift app was written by 1 psychologist, 2 clinical psychologists (JC and RC), 2 psychiatrists (SH and NG), and 1 researcher (GB).

**Phase 3: Pilot-Testing**

Pilot-testing was conducted in October and November 2019 to examine the usability and acceptability of the prototype version of the Shift app. Trends in depression and anxiety symptom severity, as well as changes in help-seeking intentions before and after using the Shift app prototype over a 4-week period were also examined. This study was approved by the South Eastern Sydney Local Health District Human Research Ethics Committee (protocol # 2019/ETH00318).

Two New South Wales hospital sites, one regional and one metropolitan, issued recruitment calls via email messages on behalf of the research team to junior physicians at the intern, resident, or registrar levels. The eligibility criteria were current employment as junior medical officers in New South Wales and ownership of an internet-enabled smartphone with an Apple or Android operating system. A total of 52 candidates accessed the study website, of which 50% (26/52) consented to participate. A final sample of 22 participants (13/22, 59% women; mean age 29, SD 4.1 years) entered the pilot study and completed a baseline assessment. A diagram of the participant flow is shown in Figure 1.
A baseline questionnaire assessed basic demographic information (age, gender, level of training, and regional or metropolitan placement) and symptoms of depression (Patient Health Questionnaire 9-item) and anxiety (General Anxiety Disorder scale 7-item) over the past 2 weeks. Participants indicated their depression and anxiety symptoms on a 4-point Likert-type response scale ranging from 0 (not at all) to 3 (nearly every day). In addition, participants indicated their previous and recent help-seeking intentions for mental health problems (eg, “If you were to develop a mental health issue, how likely would you be to seek help from a GP or mental health professional [eg, psychologist/psychiatrist]?”) on an 11-point Likert-type response scale ranging from 0 (not at all likely) to 10 (very likely). After completion of the baseline questionnaire, participants were given instructions on how to download the Shift app onto their mobile phones. The app automatically recorded usage metrics, such as the number of log-ins and challenges or sessions completed.

After 1 month, participants were invited via email to complete a poststudy questionnaire reassessing their current depression and anxiety symptom levels and help-seeking intentions. At the poststudy assessment point, participants were further asked to respond to a battery of purpose-built questions relating to the usability and acceptability of the Shift app prototype (eg, “Was the app interesting/engaging?”) on a 5-point Likert-type scale ranging from 1 (low) to 5 (high). A modified version of the System Usability Scale was also administered to obtain an objective indication of the overall ease of use of the app.

**Phase 4: Shift App Redevelopment**

It was anticipated that a final phase of design would be required after the initial pilot-testing of the Shift app. A series of 3 user-consultation workshops (N=51) at 2 metropolitan hospitals, driven by a lead user experience designer (VC), helped develop any final changes and responses to insights gained from the pilot-testing. The feedback of junior physicians participating in the workshops was collected using an interactive prototype containing screen mock-ups and questions aimed at testing potential changes and solutions.

**Results**

**Phase 1: Needs Assessment Through Qualitative Interviews**

**Attitudes Toward a Mental Health App for Junior Physicians**

All participants reported that they owned a smartphone and used apps multiple times each day. Half of the sample had already tried at least one mindfulness app, and 2 had previously used a web-based mental health app. Although the use of work-related apps was common, the use of such apps was centered on communication and medical information–seeking and reference.
material. A total of 83% (10/12) participants reported to be very or somewhat interested in the idea of an app to support the mental health of junior physicians and provided positive endorsement for the idea in principle and for its potential to benefit the population. Most of the participants endorsed use of the app for both prevention of and early intervention in mental illness and suggested the inclusion of directions to support and treatment services for those with more severe mental illness. Although no participants were directly opposed to the idea of an mHealth app to support junior physicians’ mental health, 2 participants questioned whether fellow junior physicians would use an mHealth app or to what extent they would benefit from it, particularly if physicians were not currently experiencing symptoms of mental illness (“no one takes their medicine when they are feeling good”).

### App Naming, Content, and Feature Suggestions

Shift was named based on participants’ views that the name of the app should be unobtrusive and not obviously related to mental health. This term refers to both shift work, which is one of the most common stressors in this population, and to shifting cognitions and behaviors to promote better mental health and well-being, in line with cognitive behavioral and mindfulness principles.

Participants made numerous suggestions for clinical app contents, including mindfulness, sleep strategies, mood monitoring, behavioral activation, and cognitive therapy targeting specific situations commonly faced by junior physicians. The provision of strategies to deal with problematic work situations was often seen as having a stress-buffering and destigmatizing effect. A list of recommendations for the broad content areas provided by participants is presented in Table 1.

<table>
<thead>
<tr>
<th>Recommended content components</th>
<th>Version 1</th>
<th>Version 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive behavioral therapy</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Sleep hygiene</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Mindfulness and stress management</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Goal-setting</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Pleasant activity scheduling</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Practical lifestyle strategies</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Problematic work situations</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Hand over tips for changing terms or hospitals</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Stories from junior physicians</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Version 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

*Indicates an addition or improvement compared with the previous version.

All participants provided positive feedback about the idea of a 30-day challenge, with some reporting that they were more likely to use this feature as it was time-limited and seemed achievable with a set end point and small regular goal-oriented challenges. Several participants suggested graphical feedback on symptom trackers that compared multiple outcomes over time to appeal to the scientifically minded. A list of desired mHealth app features, as expressed by the participant sample, is presented in Table 2.
Table 2. Qualitative interview participants’ desired features for Shift and whether recommendations were adopted (N=12).

<table>
<thead>
<tr>
<th>Desired features</th>
<th>Version 1</th>
<th>Version 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logical, clear app structure with clear user flow</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Simple layout, easy to navigate quickly</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Default private option—no linkage to social media</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Quick start option, easy access log in</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Skip function; ability to return to modules later</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Provision of both text and audio formats</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Symptom tracker function with graphics showing charts</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>At log in, quick tick box of symptom self-assessment</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Centralized access to many things from one place</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Notifications and reminders should be optional</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Indicates improvement compared with the previous version.

**Facilitators of and Barriers to App Use**

The main app use barrier reported by the participants was that the app would feel like another chore when participants were already time-poor and working long hours. Although most participants expressed concerns about confidentiality, deidentification, and minimal information-sharing, 92% (11/12) of participants reported that they would still provide their name and email address to register to a mental health app.

The main facilitators of app engagement that the participants reported were the app being quick and easy to use and having added-value features that would distinguish the app over others (that’s targeted to medics, that other apps aren’t going to address). Most participants reported that they would be happy to use the app quite frequently, such as every morning or every few days, but only if the sessions were very brief.

**Phase 2: Prototype Development**

The Shift app was developed for use in Android and Apple operating systems. The main features of the app are the following: (1) therapeutic and psychoeducational modules, (2) provision of contact details to mental health organizations and workplace resources, (3) mood and habit tracking, and (4) brief symptom assessments. Shift delivers content through a variety of text, audio, video, and graphical displays.

On the basis of previous research on the therapeutic benefits of cognitive behavioral, mindfulness, and value-based action components for medical professionals [14,15] and guided by preferences expressed in phase 1 qualitative interviews, the Shift app was developed employing cognitive behavioral principles of thought evaluation (ie, identification, evaluation, and modification of unhelpful thoughts) and engagement in valued action (ie, values-consistent patterns of action) adopted from acceptance and commitment therapy. Relaxation techniques (eg, progressive muscle relaxation) based on mindfulness and stress management practices were also incorporated to lessen the impact of stressful life events or daily stressors. Cognitive behavioral, value-based, and mindfulness modules were generated and presented in a 30-day challenge format. Each challenge was designed to take approximately 3-4 minutes to complete.

In addition, a suite of psychoeducational modules (sessions) was developed, including informational content on common mental health disorders, avenues through which to seek help for mental health concerns, and suggestions on how to incorporate relevant strategies, such as how to adjust to shift work. Psychoeducational content included mental health, help-seeking, and workplace information, such as depression, anxiety, mandatory reporting, at home and workplace avenues for seeking help, workplace bullying, adjusting to rural and regional placements, exams and interviews, and sleep health.

A tracking tool and symptom screening options were designed to allow users to capture daily snapshots of how they were faring and, in the case of the tracking function, build a visual tool to observe variations in mood and behavioral patterns over time.

The clinical content and design drafts of the app prototype were modified based on user experience feedback. User experience experts and psychologists worked with 2 junior physicians to refine the user pathways, functionality, color palette, and design and to modify the clinical content (eg, examples of scenarios used in cognitive therapy). Tables 3 and 4 outline the contents of the resulting Shift app prototype, which were organized into challenges and sessions.
<table>
<thead>
<tr>
<th>Therapeutic type and challenge day or days</th>
<th>Topic name</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mindfulness</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Introduction to mindfulness</td>
<td>Video</td>
</tr>
<tr>
<td>3</td>
<td>Seeing the horizon</td>
<td>Audios</td>
</tr>
<tr>
<td>9</td>
<td>Grounding anchor</td>
<td>Audios</td>
</tr>
<tr>
<td>16</td>
<td>Calming breath</td>
<td>Audios</td>
</tr>
<tr>
<td>20</td>
<td>Loving-kindness</td>
<td>Audios</td>
</tr>
<tr>
<td>24</td>
<td>Cargo thoughts</td>
<td>Audios</td>
</tr>
<tr>
<td>26</td>
<td>Breathing wind</td>
<td>Audios</td>
</tr>
<tr>
<td>28</td>
<td>Lapping ocean</td>
<td>Audios</td>
</tr>
<tr>
<td><strong>Value-based</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Introduction to values and values as a physician</td>
<td>Video</td>
</tr>
<tr>
<td>5</td>
<td>Strive for five</td>
<td>Text</td>
</tr>
<tr>
<td>6, 12, 17, 21, 27</td>
<td>Scheduling meaningful actions</td>
<td>Text</td>
</tr>
<tr>
<td><strong>Cognitive behavioral</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Introduction to unhelpful thoughts</td>
<td>Video</td>
</tr>
<tr>
<td>8</td>
<td>Unhelpful thoughts</td>
<td>Text</td>
</tr>
<tr>
<td>11</td>
<td>Cognitive biases</td>
<td>Text</td>
</tr>
<tr>
<td>14</td>
<td>Introduction to thought challenging</td>
<td>Video</td>
</tr>
<tr>
<td>15</td>
<td>Thought challenging</td>
<td>Text</td>
</tr>
<tr>
<td>22</td>
<td>Worry decision tree</td>
<td>Text</td>
</tr>
<tr>
<td>23</td>
<td>Cognitive therapy review</td>
<td>Video</td>
</tr>
<tr>
<td><strong>Positive psychology</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Gratitude</td>
<td>Text</td>
</tr>
<tr>
<td>13</td>
<td>Getting active</td>
<td>Text</td>
</tr>
<tr>
<td>18</td>
<td>Social support</td>
<td>Text</td>
</tr>
<tr>
<td>19</td>
<td>Help a friend</td>
<td>Text</td>
</tr>
<tr>
<td>25</td>
<td>100 enjoyable activities</td>
<td>Text</td>
</tr>
<tr>
<td>29</td>
<td>Planning for the future</td>
<td>Text</td>
</tr>
<tr>
<td><strong>General</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Checkup</td>
<td>Text</td>
</tr>
<tr>
<td>30</td>
<td>Putting it all together</td>
<td>Video</td>
</tr>
</tbody>
</table>
Table 4. Organization of session topics in the Shift app prototype version.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sleep and fatigue</strong></td>
<td></td>
</tr>
<tr>
<td>Sleep health</td>
<td>Text</td>
</tr>
<tr>
<td>Adjusting to shift work</td>
<td>Text</td>
</tr>
<tr>
<td><strong>Common mental health problems</strong></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>Text</td>
</tr>
<tr>
<td>Anxiety</td>
<td>Text</td>
</tr>
<tr>
<td>Burnout</td>
<td>Text</td>
</tr>
<tr>
<td>Posttraumatic stress</td>
<td>Text</td>
</tr>
<tr>
<td>Alcohol and other drugs</td>
<td>Text</td>
</tr>
<tr>
<td><strong>Getting help</strong></td>
<td></td>
</tr>
<tr>
<td>Get help now</td>
<td>Text</td>
</tr>
<tr>
<td>Dealing with intense emotions</td>
<td>Text</td>
</tr>
<tr>
<td>How to seek help: workplace avenues</td>
<td>Text</td>
</tr>
<tr>
<td>How to seek help: nonworkplace avenues</td>
<td>Text</td>
</tr>
<tr>
<td>Mandatory reporting</td>
<td>Text</td>
</tr>
<tr>
<td><strong>Common issues for JMOs</strong></td>
<td></td>
</tr>
<tr>
<td>Exams and interviews</td>
<td>Text</td>
</tr>
<tr>
<td>Work-life balance</td>
<td>Text</td>
</tr>
<tr>
<td>Adjusting to rural and regional placements</td>
<td>Text</td>
</tr>
<tr>
<td>Bullying in the workplace</td>
<td>Text</td>
</tr>
<tr>
<td>Dealing with the death of a patient</td>
<td>Text</td>
</tr>
<tr>
<td>Calling for a consult</td>
<td>Text</td>
</tr>
<tr>
<td>Feeling inadequate</td>
<td>Text</td>
</tr>
</tbody>
</table>

*aJMO: junior medical officer.*

**Phase 3: Pilot-Testing**

**App Acceptability and Usability**

As shown in Table 5, the median responses to questions relating to overall app rating, content understandability, appropriateness, and usefulness were all on or above the midrange of the response scales. The overall system usability rating was 84.72 (SD 8.33), which was above the average score of 70 across technological tools more generally [37] and comparable with an average score of 77 reported for common smartphone apps and tablets [38]. Tables 6 and 7 present a breakdown of participants’ usefulness ratings of Shift challenge and session components. Participants rated the mindfulness challenges favorably and rated the general and value-based components least favorably. Among the session topics, the sleep and fatigue information components received the highest usefulness ratings, while the other sessions were rated lower or were not attempted.

Table 5. Pilot trial participants’ responses to the usability and acceptability of the Shift app prototype (N=9)*.

<table>
<thead>
<tr>
<th>Item</th>
<th>Values, median (range)</th>
<th>Values, minimum-maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>How well did you understand the content of the app?</td>
<td>5 (2)</td>
<td>3-5</td>
</tr>
<tr>
<td>Was the app content appropriate for you?</td>
<td>5 (3)</td>
<td>2-5</td>
</tr>
<tr>
<td>Was the app interesting/engaging?</td>
<td>4 (2)</td>
<td>3-5</td>
</tr>
<tr>
<td>Do you feel that the app has helped you improve your mental health?</td>
<td>3 (2)</td>
<td>1-3</td>
</tr>
<tr>
<td>Would you recommend this app to other junior physicians?</td>
<td>4 (2)</td>
<td>3-5</td>
</tr>
<tr>
<td>What is your overall rating of the app?</td>
<td>3 (3)</td>
<td>2-5</td>
</tr>
</tbody>
</table>

*aResponse scales ranged from 1 to 5.*
### Table 6. Pilot trial participants’ usefulness ratings of the Shift app prototype 30-day challenge contents (N=8)\(^a\).

<table>
<thead>
<tr>
<th>Challenge type and topic name(^b)</th>
<th>Values, median (range)</th>
<th>Values, minimum-maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mindfulness</td>
<td>4 (5)</td>
<td>0-5</td>
</tr>
<tr>
<td>Value-based</td>
<td>0 (5)</td>
<td>0-5</td>
</tr>
<tr>
<td>Cognitive behavioral</td>
<td>3 (5)</td>
<td>0-5</td>
</tr>
<tr>
<td><strong>Positive psychology</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gratitude</td>
<td>3 (5)</td>
<td>0-5</td>
</tr>
<tr>
<td>Getting active</td>
<td>1.5 (5)</td>
<td>0-5</td>
</tr>
<tr>
<td>Social support</td>
<td>2.5 (5)</td>
<td>0-5</td>
</tr>
<tr>
<td>Help a friend</td>
<td>0 (4)</td>
<td>0-4</td>
</tr>
<tr>
<td>Enjoyable activities</td>
<td>3 (5)</td>
<td>0-5</td>
</tr>
<tr>
<td>Planning for the future</td>
<td>1.5 (5)</td>
<td>0-5</td>
</tr>
<tr>
<td>General</td>
<td>0 (4)</td>
<td>0-4</td>
</tr>
</tbody>
</table>

\(^a\)Response scales ranged from 0 to 5, where 0 indicates unattempted components, 1 indicates low perceived usefulness, and 5 indicates high perceived usefulness.

\(^b\)Only positive psychology challenge topics were assessed individually because of the distinctiveness of each topic in this category.

### Table 7. Pilot trial participants’ usefulness ratings of the Shift app prototype session contents (N=8)\(^a\).

<table>
<thead>
<tr>
<th>Session type and topic name</th>
<th>Values, median (range)</th>
<th>Values, minimum-maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sleep and fatigue</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sleep health</td>
<td>3 (5)</td>
<td>0-5</td>
</tr>
<tr>
<td>Adjusting to shift work</td>
<td>3 (5)</td>
<td>0-5</td>
</tr>
<tr>
<td><strong>Common mental health problems</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>1.5 (4)</td>
<td>0-4</td>
</tr>
<tr>
<td>Anxiety</td>
<td>2.5 (5)</td>
<td>0-5</td>
</tr>
<tr>
<td>Burnout</td>
<td>1.5 (5)</td>
<td>0-5</td>
</tr>
<tr>
<td>Posttraumatic stress</td>
<td>0 (4)</td>
<td>0-4</td>
</tr>
<tr>
<td>Alcohol and other drugs</td>
<td>0 (4)</td>
<td>0-4</td>
</tr>
<tr>
<td><strong>Getting help</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intense emotions</td>
<td>0 (4)</td>
<td>0-4</td>
</tr>
<tr>
<td>Workplace avenues</td>
<td>1 (5)</td>
<td>0-5</td>
</tr>
<tr>
<td>Nonworkplace avenues</td>
<td>1.5 (5)</td>
<td>0-5</td>
</tr>
<tr>
<td>Mandatory reporting</td>
<td>0 (5)</td>
<td>0-5</td>
</tr>
<tr>
<td><strong>Common issues for JMOs(^b)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exams and interviews</td>
<td>0 (3)</td>
<td>0-3</td>
</tr>
<tr>
<td>Work-life balance</td>
<td>1.5 (5)</td>
<td>0-5</td>
</tr>
<tr>
<td>Remote placements</td>
<td>0 (4)</td>
<td>0-4</td>
</tr>
<tr>
<td>Workplace bullying</td>
<td>0 (5)</td>
<td>0-5</td>
</tr>
<tr>
<td>Death of a patient</td>
<td>0 (5)</td>
<td>0-5</td>
</tr>
</tbody>
</table>

\(^a\)Response scales ranged from 0 to 5, where 0 indicates unattempted components, 1 indicates low perceived usefulness, and 5 indicates high perceived usefulness.

\(^b\)JMO: junior medical officer.
**App Use**

Of the 22 participants, 95% (21) downloaded *Shift* and used the app at least once. The mean number of log-ins was 5.24 (SD 5.93) and participants spent an average of 28 minutes (SD 52.7) on the app, although the large SD indicated that there was considerable variability in use times. The median use time was 11 minutes (range 199.35). Participants completed an average of 5 challenges (SD 7.75) and spent 3.24 minutes working through these challenges (SD 4.53). Sessions were used less frequently (mean 1.81, SD 2.56), and less time was spent on sessions (mean 1.38, SD 1.46 minutes).

**Symptoms Change**

Wilcoxon signed-rank tests indicated that depression ($Z=-1.38; P=.168$) and anxiety ($Z=-1.05; P=.293$) scores slightly decreased, albeit not significantly, and that help-seeking intentions were largely unchanged ($Z=-.38; P=.705$) over the 1-month period of app use.

**Phase 4: Shift App Redevelopment**

On the basis of the results of the pilot, a major redesign of the prototype was conducted to create a more user-friendly and user-driven learning experience. The pilot study results indicated that the first version of the app did not engage junior physicians sufficiently well, which may have been an important factor contributing to small effect sizes in symptom change and help-seeking scores. A series of 3 user-consultation workshops (N=51) at 2 metropolitan hospitals helped finalize the proposed changes to the prototype. These were (1) layout and design improvements, (2) increased personalization and ease of access, (3) updates of clinical contents, (4) the inclusion of self-reflection activities, and (5) the adoption of more meaningful and relatable wording [6].

A streamlined login process with the inclusion of a biometric security system (ie, fingerprint or face ID authentication) was incorporated to facilitate use after app download. Personalization enhancements were achieved through significant changes in the presentation of the app. Importantly, the 30-day challenge structure was removed, and challenges and sessions were organized under a general overview of the topics. In this view, users were directed to contents through headings named *Mental Health, Getting Help, Lifestyle, and Work*. The challenge concept was maintained by incorporating an option for users to set their own weekly targets (ie, number of activities to complete each week) during the app on boarding process. With this new functionality, users could choose achievable goals relating to their app use while still encouraging regular use of the app.

The previously limited line-graph tracking function was completely redesigned to become more interactive and visual, as well as allowing a new option of *work/life balance*. To accommodate these changes, thorough layout and presentation updates were made, as well as enhancing the interactivity with a complimentary day-by-day view to show which activities were completed on which dates. Consistent with the user experience workshop feedback, new modules on exercise and diet were incorporated into the app. Two additional modules were developed in response to the COVID-19 pandemic. Informational sessions were extended with the inclusion of example stories from junior physicians and adjunct brief empiric evaluations (*symptom screeners*, for example, the Patient Health Questionnaire–2-item, General Anxiety Disorder scale–2-item). Example stories, based on recommendations put forward by junior physicians in qualitative interviews (Table 1), illustrated mental health challenges and invited users to elaborate on symptoms. The mHealth app contents were finally proofread by 2 clinical psychologists, 3 researchers (IC, AB, and SS), and 1 digital learning designer (JP) to ensure the provision of up-to-date clinical and psychoeducational contents in a way that facilitates learning [39,40]. The structure and design graphical guides of the novel *Shift* version are provided in Figure 2.
Figure 2. Visual examples of the most recent *Shift* version Home, Topics, Tracking, and Settings screens.

**Discussion**

**Principal Findings**

This paper describes the development of an mHealth app, *Shift*, designed to support the mental health and help-seeking of junior physicians. In line with gold standard recommendations on the importance of user-centered design principles [26], the 4-phased app development process (ie, semistructured interviews, prototype development, pilot-testing, and app redevelopment) focused on a participatory approach to promote effective engagement and facilitate cognitive, affective, and behavior changes of junior physicians. Junior physicians were involved at every stage of this process through qualitative interviews, user experience workshops, and participation in pilot-testing. The aim of this approach was to create, deliver, and refine content in a way that was acceptable, effective, and engaging to end users.

Pilot-testing revealed several issues with the delivery of in-app content components to junior physicians. Although a successive, day-by-day delivery of therapeutic content has been successfully employed in a previous working population sample [21] and was generally viewed favorably by junior physicians in qualitative assessments, preliminary use data indicated that this format failed to engage junior physicians in practice. Inspection of app use data revealed that discontinuation of the 30-day challenge tended to appear around day 5, which was a generally lower rated, value-based activity. With the 30-day challenge format, users were unable to skip challenge topics or change the order of challenges, possibly facilitating the discontinuation of app use. As a time-poor, well-educated group, junior physicians may be more insistent on being able to choose their own modules from other working populations. In addition, informational sessions were underutilized in comparison to challenge content, possibly due to their less-prominent positioning within the app. Feedback on challenge and session components indicated considerable variability in the favorability ratings of the contents. This observation highlights 2 key aspects. First, even when app components are generated based on qualitative data from focus groups, their use in the real world needs to be tested. Second, simply modifying the modes of content delivery (ie, 30-day challenge structure) from one evidence-based app to suit another working population is not always successful. To meet the needs of a population of junior physicians, a new app structure needs to be developed. Therefore, in phase 4, the app was adjusted with the help of junior physicians across successive user experience workshops to enhance the overall experience and encourage engagement. The main changes included streamlining login and onboarding procedures and categorizing contents by topics, which allowed for the personal selection of modules and for an updated design and learning experience.

**Strengths and Limitations**

The 4-phase process emphasizes the need for customization for end users. In line with previous research, this project illustrates the role of usability testing in the development of a digital intervention tool [24,41]. Using participatory mixed methods, such as qualitative and quantitative assessments, to involve end users at all stages of the product development process was fundamental in our attempt to create a digital solution that allowed for the pursuit of multiple outcomes, such as cognitive behavioral, psychoeducational, or providing contact details to relevant specialized services [26]. Furthermore, our development process suggests that user experience and learning designers are critical in translating methods proven useful in face-to-face settings in the digital arena. The establishment of a multidisciplinary team including academics, clinicians, and digital experts helped incorporate the suggestions and feedback...
put forward by junior physicians into an mHealth app intervention environment.

This study had several important limitations. First, the Shift mHealth app has not been evaluated for its effectiveness to date. Future studies are required to establish whether it is indeed a useful tool to reduce or prevent the onset of common mental health symptoms in junior physicians. Similarly, it needs to be established whether help-seeking intentions or actions improve after using Shift. Second, although the increased personalization of the novel Shift version is expected to increase engagement, the freedom to choose modules may unintentionally facilitate avoidance behaviors or choice overload and thus potentially minimize exposure to beneficial content. Thus, use behaviors and outcomes should be thoroughly inspected to predict which use patterns constitute effective engagement if such a pathway indeed exists. Finally, Shift focuses on individual-level change factors. However, to address distress among junior physicians in a comprehensive manner, structural change programs including organization-level solutions are likely to be required alongside the delivery of interventions directed at an individual, such as the app [14].

Conclusions

The integration of new technology in the creation of workplace and personal well-being programs constitutes an easily accessible and cost-effective approach to addressing mental health concerns. Digital mental health programs provide a potential solution to engaging hard-to-reach, time-poor, and potentially help-seeking averse junior physicians in a way that does not exacerbate confidentiality concerns around discussing mental health problems face-to-face. To our knowledge, Shift is the first initiative of its kind in that it aims to deliver mental health and support resources to junior physicians through unobtrusive mHealth app technology. This study describes an innovative, multiphase, and multidisciplinary user-driven design process that was undertaken to ensure a match between the app and the needs and barriers faced by junior physicians. Further research is planned to examine whether Shift proves useful for a substantial number of junior physicians.

Acknowledgments

The authors would like to thank Jennifer Chapman and Deborah Frew from the New South Wales Ministry of Health and the Shift Steering Committee for their continued support and guidance on all aspects related to the development and promotion of Shift. They are also extremely grateful to all their contacts across New South Wales’ Local Health Districts and hospitals for promoting the Shift study across their channels, to all junior physicians who volunteered to participate in the development and testing of the Shift app, and to the Black Dog Institute IT Department for their ongoing information technology and development support.

Conflicts of Interest

The development of Shift was made possible with funding from the New South Wales Ministry of Health and icare Foundation. All researchers have remained independent from the funders in the completion and submission of this work. Intellectual property for the Shift app is owned by the Black Dog Institute; however, the Shift app does not currently produce any income and the authors do not receive any financial gain from this intellectual property.

References


27. Counson et al JMIR FORMATIVE RESEARCH [Medline: 32257673]


Abbreviations

mHealth: mobile health

©Isabelle Counson, Alexandra Bartholomew, Joanna Crawford, Katherine Petrie, Geetanjali Basarkod, Victoria Moynihan, Josie Pires, Rachel Cohen, Nicholas Glozier, Samuel Harvey, Samineh Sanatkar. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
The Use of Food Images and Crowdsourcing to Capture Real-time Eating Behaviors: Acceptability and Usability Study

Katharine Harrington¹, MPH; Shannon N Zenk²,³, PhD; Linda Van Horn¹, PhD; Lauren Giurini⁴; Nithya Mahakala¹; Kiarri N Kershaw¹, PhD

¹Northwestern University Feinberg School of Medicine, Chicago, IL, United States
²National Institute of Nursing Research, Bethesda, MD, United States
³National Institute on Minority Health and Health Disparities, Bethesda, MD, United States
⁴Lake Forest College, Lake Forest, IL, United States

Corresponding Author:
Kiarri N Kershaw, PhD
Northwestern University Feinberg School of Medicine
680 N Lake Shore, Suite 1400
Chicago, IL, 60611
United States
Phone: 1 312 503 4014
Email: k-kershaw@northwestern.edu

Abstract

Background: As poor diet quality is a significant risk factor for multiple noncommunicable diseases prevalent in the United States, it is important that methods be developed to accurately capture eating behavior data. There is growing interest in the use of ecological momentary assessments to collect data on health behaviors and their predictors on a micro timescale (at different points within or across days); however, documenting eating behaviors remains a challenge.

Objective: This pilot study (N=48) aims to examine the feasibility—usability and acceptability—of using smartphone-captured and crowdsource-labeled images to document eating behaviors in real time.

Methods: Participants completed the Block Fat/Sugar/Fruit/Vegetable Screener to provide a measure of their typical eating behavior, then took pictures of their meals and snacks and answered brief survey questions for 7 consecutive days using a commercially available smartphone app. Participant acceptability was determined through a questionnaire regarding their experiences administered at the end of the study. The images of meals and snacks were uploaded to Amazon Mechanical Turk (MTurk), a crowdsourcing distributed human intelligence platform, where 2 Workers assigned a count of food categories to the images (fruits, vegetables, salty snacks, and sweet snacks). The agreement among MTurk Workers was assessed, and weekly food counts were calculated and compared with the Screener responses.

Results: Participants reported little difficulty in uploading photographs and remembered to take photographs most of the time. Crowdsource-labeled images (n=1014) showed moderate agreement between the MTurk Worker responses for vegetables (688/1014, 67.85%) and high agreement for all other food categories (871/1014, 85.89% for fruits; 847/1014, 83.53% for salty snacks, and 833/1014, 81.15% for sweet snacks). There were no significant differences in weekly food consumption between the food images and the Block Screener, suggesting that this approach may measure typical eating behaviors as accurately as traditional methods, with lesser burden on participants.

Conclusions: Our approach offers a potentially time-efficient and cost-effective strategy for capturing eating events in real time.

(JMIR Form Res 2021;5(12):e27512) doi:10.2196/27512

KEYWORDS
ecological momentary assessment; eating behaviors; crowdsourcing; food consumption images; food image processing; mobile phone
Introduction

Background

Poor diet quality is a significant risk factor for multiple noncommunicable diseases, including diabetes, certain cancers, and cardiovascular disease [1-3]; however, effective strategies for promoting healthful dietary behavior changes remain elusive. Data reported by the American Heart Association show that <2% of US adults consume an ideal diet [4], a finding further supported by similar data indicating suboptimal intake in most countries [5]. Changing eating behaviors is challenging, partly because of the multifactorial influences on eating decisions. These range from individual and family-level beliefs, preferences, and constraints to larger social, physical, environmental, and temporal and situational cues [6-9]. Such complexity surrounding eating decisions suggests the importance of documenting not only what people eat but also the social and contextual factors potentially influencing their choices.

A growing number of studies have used ecological momentary assessments (EMAs) to simultaneously capture information on what people eat and the role that various social and contextual factors play in influencing those decisions on a micro timescale (across eating events over 1 or several days) [10]. EMAs involve repeatedly sampling participants’ behaviors and experiences in real time within their natural environments [11]. This typically involves administering surveys several times throughout the day using SMS text messaging or a smartphone app. EMA has been used in several studies to evaluate the predictors of intraindividual changes in eating behaviors throughout the day or across days [12-14]. This timescale and the widespread use of smartphones simplify the evaluation across a wide range of predictors, including stress, social and physical environments, and time of day.

To confidently determine the predictors of eating behaviors, we need to accurately capture eating events. Measuring eating behaviors on a micro timescale makes some of the more traditional self-reported dietary measures less practical or useful [15,16]. For example, a 24-hour dietary recall is difficult and burdensome for participants to document through SMS text messaging or a smartphone app. In addition, this format of data collection would require participants to recall their emotions at the time of the meals or provide further details regarding their environment during each meal, which could lead to measurement error and recall bias. Image-based food data collection methods have been developed and evaluated for measuring energy intake; however, they often require participants to use a fiducial marker when taking the images, followed by time-intensive analyses by a dietitian [17,18]. This approach is useful when quantifying total energy intake or when nutrients are of central importance but less so for measuring the predictors of fluctuations in eating behaviors throughout the day or from day to day (eg, snacking or unhealthy eating events).

Most EMA studies seeking to measure the predictors of eating behavior on a micro timescale require that participants record their eating events through diaries or journal entries [19,20] or through the completion of checklists having a variety of different food types [21-24]. This is problematic, as such lists are finite and may fail to fully capture the relevant food options, especially on a smartphone screen. In addition, the act of checking a box to confirm certain eating decisions may influence and alter behaviors [25]. Thus, an alternative approach that is more comprehensive but less overt is warranted.

Objective

This study assesses the feasibility—the ease of use and participant acceptability—of coupling participant-captured images with crowdsourcing to document eating events in real time. Collecting images facilitates and enhances the self-reported measures of food consumption. Photographing food is now commonplace and socially acceptable, thereby offering a practical strategy for obtaining comprehensive assessments of eating behaviors while lessening the burden on participants. Furthermore, the use of crowdsourcing to classify and quantify food items is a time- and cost-effective, scalable approach with proven accuracy in other biomedical applications [26]. Crowdsourcing minimizes participant burden by eliminating the need to label food images themselves or fill out dietary recall journals and surveys. Through this elimination, crowdsourcing also limits participant reflection on their eating decisions, which could alter their behaviors during the study time frame. Implemented together, participant-captured images and crowdsourcing of image labels can provide a feasible alternative to current food intake EMA methods.

Methods

Study Population

We recruited a convenience sample of 48 former participants of the Chicago Healthy Eating Environments and Resources Study (CHEERS) to participate in this pilot study. CHEERS was a cross-sectional study of 228 nonpregnant women aged 18-44 years and living in 4 racially, ethnically, and socioeconomically diverse neighborhoods in Chicago, Illinois, who could understand English or Spanish [27]. The women were recruited via flyers posted in neighborhood stores, presentations to parent organizations at schools, and mails sent using commercially available address lists. This study focused on recruiting women because of the persistent racial and ethnic obesity disparities that exist among women and as women are typically responsible for food preparation and purchase in their families. Women in this age group were selected, as these years are a critical period for an increase in weight because of a range of factors, including postpregnancy weight retention and declining muscle mass and muscle strength [28]. Data for the original study were collected between 2016 and 2017, whereas data for this pilot study were collected between 2018 and 2019. This study was approved by the Northwestern University institutional review board (STU00203035), and all participants gave informed consent.

Study Procedure

The CHEERS EMA pilot study comprised a 1-day initial visit, a 7-day EMA study period, and a 1-day final visit. Participants were incentivized to participate with cash rewards: US $20 for the first visit, US $9 per day of valid data collection (up to US $63), and US $30 for the final visit. During the initial visit,
participants were asked to complete several questionnaires adopted from other studies or created specifically for CHEERS. Participants also completed the validated Block Fat/Sugar/Fruit/Vegetable Screener, which provides estimates of saturated fat, trans fat, total sugars, added sugars (in sweetened cereals, soft drinks, and sweets), fruit and fruit juice, vegetable intake, glycemic load, and glycemic index [29]. In this study, the Block Fat/Sugar/Fruit/Vegetable Screener was used to estimate the usual weekly intake of fruits, vegetables, sweet snacks, and salty snacks.

During the 7-day EMA study period, participants documented all their meals and snacks by capturing images via smartphones using 1 of 2 apps: LifeData (LifeData, LLC) or Mobile EMA (ilumivu, Inc). During the initial visit, study team members installed the apps either on participants' personal smartphones or on a study-provided smartphone, and participants were trained to use the apps. Study data were stored within the app and then uploaded to the server when connected to Wi-Fi during the study period or when the phone was returned at the end of data collection; therefore, Wi-Fi was not required for data collection.

A combination of event- and signal-contingent EMA surveys were used. For the event-contingent surveys, participants answered 3 short questions at the time each food item image was uploaded. Participants were asked to upload a picture of their meal or snack at the time of each eating episode, and they received an SMS text message every morning reminding them to complete the meal or snack event-contingent surveys. Participants indicated whether the food item was a meal or snack. Trained staff contacted the participants on day 2 of the 7-day data collection period to answer questions and encourage adherence. Signal-contingent surveys were randomly sent throughout the day to assess the role of stress and affect on eating behaviors. Participants were asked to complete 4 surveys per day. They received prompts between 8 AM and 8 PM, with at least 2 hours between prompts. Participants’ phones were set to allow push notifications to alert them as the prompts came through. Each survey was available for 10 minutes to more accurately capture stress and affect in real time. If the meal or snack was not photographed, participants were requested to write down what they ate and upload an image of that description. During the study, participants also wore heart rate monitors and accelerometers to assess their physical activity and stress levels. Participants were rewarded US $4 for each day with valid heart rate and accelerometer data and US $5 for each day with at least 3 signal-contingent and 2 event-contingent EMA surveys.

**Usual Food Intake**

Participants completed the Block Fat/Sugar/Fruit/Vegetable Screener [29] to assess their usual intake of foods relevant to the study. The Block Fat/Sugar/Fruit/Vegetable Screener (Screener) is a food frequency questionnaire that has been validated for providing estimates of saturated fat, trans fat, total sugar, fruits, and vegetables. Food frequencies were determined through participant responses to “How many days per week?” for the relevant survey questions (Textbox 1). Participants were asked to select either none or less than 1, 1 day, 2 days, 3-4 days, 5-6 days, or every day. Responses were coded as 0, 1, 2, 3.5, 5.5, and 7 and summed by food category. This measure was used to reflect the number of times per week an individual usually consumed that particular food group, and it was compared with the number of times calculated from the images they submitted. As with Amazon Mechanical Turk (MTurk)–processed images, the Screener aimed to capture the number of times each type of food was eaten, not the serving size.
Textbox 1. Questions taken from the Block Fat/Sugar/Fruit/Vegetable Screener at previsit to determine the self-reported intake of fruit, vegetables, salty snacks, and sweet snacks.

<table>
<thead>
<tr>
<th>Fruit</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Any kind of fruit, fresh or canned (not counting juice)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vegetables</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Green salad and vegetables you put in green salad</td>
</tr>
<tr>
<td>• Potatoes, not fried, such as baked or mashed</td>
</tr>
<tr>
<td>• Vegetable soup or stew with vegetables</td>
</tr>
<tr>
<td>• All other vegetables you eat as a side dish or in any kind of dish, not counting salad or potatoes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Salty snacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>• French fries, home fries, and hash browns</td>
</tr>
<tr>
<td>• Snack chips such as potato chips, tortilla, corn chips, Fritos, Doritos, or popcorn (not pretzels)</td>
</tr>
<tr>
<td>• Crackers such Ritz, soda crackers, Cheez-Its, or any other snack cracker</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sweet snacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Ice cream and ice cream bars</td>
</tr>
<tr>
<td>• Donuts</td>
</tr>
<tr>
<td>• Cake, cookies, or snack cakes such as cupcakes, Twinkies, or any other pastry</td>
</tr>
<tr>
<td>• Pie, including fast food pies or snack pies</td>
</tr>
<tr>
<td>• Chocolate candy such as chocolate bars, M&amp;Ms, Mars Bars, and Reeses</td>
</tr>
<tr>
<td>• Any other candy (not chocolate) such as hard candy, Lifesavers, Skittles, and Starburst</td>
</tr>
</tbody>
</table>

Acceptability of Using EMA to Capture Meals and Snacks and Data Quality

After the EMA study period, participants were asked to rate their experiences with the process of taking pictures of their meals and snacks. Specifically, on 5-point Likert scales, participants were asked how often they remembered to take and upload pictures of their meals and snacks (response options included none of the time [value=1], some of the time, half of the time, most of the time, or all of the time [value=5]), how much taking the pictures changed their eating behaviors (not at all=1, slightly, somewhat, moderately, or substantially=5), whether they had difficulty uploading pictures of their food (strongly disagree=1, disagree, neither agree nor disagree, agree, or strongly agree=5), whether they had difficulty understanding the questions (strongly disagree=1, disagree, neither agree nor disagree, agree, or strongly agree=5), and whether they had difficulty entering their responses (strongly disagree=1, disagree, neither agree nor disagree, agree, or strongly agree=5). Participant acceptability questions ranged from strongly disagree (value of 1) to strongly agree (value of 5); therefore, in these questions, a higher score correlated with a higher degree of difficulty for that topic.

Crowdsourced Labeling of Food Images

MTurk was used to process the images of participants’ meals and snacks. MTurk is a crowdsourcing distributed human intelligence platform that has been used to process images for biomedical research [30-32]. Users upload discrete human intelligence tasks (HITs) that Workers can complete quickly for a small payment. In this study, 1 HIT required assigning the number of a particular food item in an image. Workers receive feedback on their performance through user approval or rejection of the HITs. Users can also specify the Worker qualifications to improve the quality of their responses.

For this study, Workers were required to have >1000 approved HITs, with an approval rate of ≥99%, and they had to be located in the United States. Eligible Workers are randomly assigned to HITs and can complete as many as they choose, resulting in many Workers completing an assignment. In this study, Workers were asked to assign counts of the following food categories in separate HITs: fruits, vegetables, salty calorie-dense foods (eg, potato chips and fries), and sweet calorie-dense foods (eg, cake, cookies, ice cream, candy, chocolate, and other pastries). Workers only counted the different food items within the current category, and they were not asked to quantify by serving size. For example, if the category was fruit and the image contained 2 grapes and half an apple, the count assigned would be 2, although it may not be equivalent to 2 servings of fruit. This is consistent with the type of eating behavior data collected in other EMA studies [10,14,33]. Screenshots of the instructions provided to Workers can be found in Figure 1.
Figure 1 displays the interface that MTurk Workers were presented. Along with the prompt, “How many different types of [food category] appear in this image?”, the Workers were provided with examples of the food in question and detailed instructions. Workers could select 0, 1, 2, 3, 4, 5+, or uncertain to categorize the image (Figure 1). If uncertain was selected, the Worker was asked to elaborate in the space provided.

**Participant Image–Captured Food Frequency**

Responses from the MTurk Workers were downloaded as comma-separated value files in the same batches as the images were uploaded. Files were combined and cleaned to ensure the absence of duplicates. As 2 Workers were assigned to count the food items for each image, their responses were compared.

Before uploading the images, a study team member checked each image to ensure that they did not accidentally include the participants’ personal information or anything that could potentially identify participants. Each batch contained approximately 100 images (HITs), and although participants were given 1 hour to complete each HIT (ie, assign the number of food items in 1 image), on average, whole batches were completed in 1 hour and 45 minutes (SD 69 minutes). Batches can be run simultaneously; therefore, all images could be processed within the same 2-hour period. Each HIT was estimated to take the Workers 15 seconds to complete, and they were provided a US $0.05 reward upon completion of each HIT. This is equivalent to US $13 per hour, which was the minimum wage in the city of Chicago at the time of the study.
Images with discordant responses between Workers were evaluated by a study team member and given a final count. Weekly intake was calculated for each participant by summing the items across the images by participant ID for each food category (fruits, vegetables, salty snacks, and sweet snacks) using the final counts assigned to the images.

**Statistical Analysis**

Acceptability of photographing food and uploading the images was evaluated based on responses to survey questions and the total number of photographs taken by participants over the EMA study period. Survey responses were categorized on Likert scales, and mean values were calculated for each component across participants. The feasibility of using MTurk to count the number of fruits, vegetables, sweet snacks, and salty snacks for each participant was assessed by calculating the percentage agreement between the responses provided by the 2 MTurk Workers. Qualitative analyses were performed to better understand the reasons for any discordance between Workers; specifically, the image was assessed for the likely reason behind the discordance, the reasons were grouped into common themes, and the frequency of each theme was calculated separately for each food category. Weekly mean food intake by category was calculated for the food images and the Screener responses. Mean values were compared using 2-tailed t tests.

**Results**

**Demographics**

Table 1 lists the study participants’ sociodemographic characteristics. The average age was 37.5 years. Of the 48 participants, 23 (48%) identified as non-Hispanic White, 5 (10%) as non-Hispanic Black, and 19 (40%) as Hispanic or Latina. Approximately 68% (32/48) of participants had at least a bachelor’s degree or higher, and 44% (21/48) found it somewhat hard or harder to pay for the basics. All participants were female.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
</tr>
<tr>
<td>&lt;40</td>
<td>30 (63)</td>
</tr>
<tr>
<td>≥40</td>
<td>18 (38)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>23 (48)</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>5 (10)</td>
</tr>
<tr>
<td>Hispanic or Latina</td>
<td>19 (40)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
</tr>
<tr>
<td>Less than a bachelor’s degree</td>
<td>15 (32)</td>
</tr>
<tr>
<td>Bachelor’s degree or higher</td>
<td>32 (68)</td>
</tr>
<tr>
<td>Financial burden</td>
<td></td>
</tr>
<tr>
<td>Not very hard to pay for basics</td>
<td>27 (56)</td>
</tr>
<tr>
<td>Hard to pay for basics</td>
<td>21 (44)</td>
</tr>
</tbody>
</table>

*aOne participant was missing demographic information.

**Acceptability and Data Quality**

A total of 1022 images were collected by the participants. Approximately 3.03% (31/1022) of images contained a written description of what was eaten in English, and 0.78% (8/1022) of images contained a description written in Spanish. Images containing a description in English were included in the batches uploaded to MTurk, and the counts were assigned by the Workers. Images with Spanish descriptions were translated, and a trained study team member assigned the counts of fruits, vegetables, salty snacks, and sweet snacks. These images were not uploaded to MTurk but were included in the final analyses comparing image and Screener responses.

Table 2 provides the average survey responses. Participants reported no difficulty in entering responses (mean 1.40, SD 0.71), understanding the questions (mean 1.48, SD 0.90), or uploading photographs (mean 2.15, SD 1.24). An average of 21.3 photographs (SD 9.52) per participant were taken over the study period. Participants remembered to take photographs more than half of the time on average, and their behavior changed slightly to somewhat because of participation in the study.
Table 2. Responses to survey questions on participant acceptability and data quality.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Score, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participant acceptability</strong>&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Difficulty entering responses</td>
<td>1.40 (0.71)</td>
</tr>
<tr>
<td>Difficulty understanding questions</td>
<td>1.48 (0.90)</td>
</tr>
<tr>
<td>Difficulty uploading photographs</td>
<td>2.15 (1.24)</td>
</tr>
<tr>
<td><strong>Data quality</strong></td>
<td></td>
</tr>
<tr>
<td>Number of photographs taken</td>
<td>21.29 (9.52)</td>
</tr>
<tr>
<td>Remembered to take photographs&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3.85 (0.87)</td>
</tr>
<tr>
<td>How much did taking pictures change behavior&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2.38 (1.18)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Possible options for each question in this section ranged from strongly disagree (1) to strongly agree (5).

<sup>b</sup>Possible options ranged from none of the time (1) to all of the time (5).

<sup>c</sup>Possible options ranged from not at all (1) to substantially (5).

**Feasibility of Using MTurk**

After the 7-day study period, 99.22% (1014/1022) of photographs of participants’ meals and snacks were assessed by MTurk Workers. Each image was uploaded and evaluated for the presence of fruits, vegetables, salty snacks, and sweet snacks; therefore, 4056 HITs were completed by Workers. On average, the batches took 1 hour and 45 minutes to process. Classification agreement among the MTurk Workers was moderate for vegetables (688/1014, 67.85%; images received the same response from both Workers), and agreement was high for all other food categories (871/1014, 85.89% for fruits; 847/1014, 83.53% for salty snacks; and 833/1014, 82.15% for sweet snacks; images received the same response). The study team performed a thematic analysis of the images that received discordant responses. A total of 4 categories were identified that presented possible explanations for the discordance: poor image quality, image subject uncertainty, user error, and miscellaneous. Images in the poor image quality category were blurry, had low visibility or a dark contrast, or the background was confusing or misleading. The image subject uncertainty category included foods that may have been difficult to discern or that portrayed a mixture of several items, such as salads, rice bowls, or vegetable mixes. In the user error category, image answers were in contrast to the provided MTurk instructions (ie, Workers were instructed to count tomatoes as its nutritional designation, as a vegetable, despite botanically being classified as a fruit), or the Worker inaccurately counted the types of food in question. The miscellaneous category applied to images that failed to fit into these 3 main classes.

**Figure 2** presents the prevalence of the 4 types of discordant response explanations by food category. User error was the most prevalent reason for discordant responses across all food categories: it was most prevalent in salty snacks (145/167, 86.8% of images with discordant responses), followed by sweet snacks (145/181, 80.1%), fruits (78/143, 54.5%), and finally vegetables (163/326, 50%). **Figure 3** further breaks down user errors into its subgroups: incorrect responses despite instruction clarification and incorrect responses because of Worker inaccuracies not specifically addressed in the instructions. Within these groups, the latter was most prevalent across all food types, with the highest percentage in salty snacks (135/145, 93.1% of images) and the lowest percentage in fruits (41/78, 53% of images).

**Figure 2.** Prevalence of discordant Mechanical Turk Worker response explanations by food category.
Image subject uncertainty was the next most common explanation for discordant responses. This explained 48.8% (159/326) of vegetable discordances, 39.2% (56/143) of fruit discordances, 17.7% (32/181) of sweet snacks discordances, and 13.2% (22/167) of salty snacks discordances (Figure 2). Image quality explained 1.2% (4/326) of vegetables, 4.9% (7/143) of fruit, 0% (0/167) of salty snacks, and 1.1% (2/181) of sweet snacks discordances. Finally, only 0.39% (4/1014) of responses were grouped into the miscellaneous category: 50% (2/4) in the fruit and 50% (2/4) in the sweet snacks responses.

Comparison of Usual Consumption From Food Images and Dietary Screener

The food images and the Screener were able to capture similar patterns of food intake. In both methods, vegetables were reportedly consumed most frequently and fruits least frequently (Table 3). Although the frequency of sweet snack consumption was lower using food images compared with that using the Screener, there were no statistically significant differences in the weekly frequency of food consumption between the image and the Screener results across all 4 food categories (fruit, $P=.99$; vegetable, $P=.54$; salty snacks, $P=.56$; and sweet snacks, $P=.37$).

Table 3. Weekly food consumption (times per week) based on food images and the Block Screener.

<table>
<thead>
<tr>
<th>Food category</th>
<th>Frequency, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fruit</td>
<td></td>
</tr>
<tr>
<td>Images</td>
<td>3.58 (4.20)</td>
</tr>
<tr>
<td>Block Screener</td>
<td>3.59 (2.32)</td>
</tr>
<tr>
<td>Vegetables</td>
<td></td>
</tr>
<tr>
<td>Images</td>
<td>9.96 (6.01)</td>
</tr>
<tr>
<td>Block Screener</td>
<td>10.28 (4.58)</td>
</tr>
<tr>
<td>Salty snacks</td>
<td></td>
</tr>
<tr>
<td>Images</td>
<td>4.17 (3.21)</td>
</tr>
<tr>
<td>Block Screener</td>
<td>3.85 (3.16)</td>
</tr>
<tr>
<td>Sweet snacks</td>
<td></td>
</tr>
<tr>
<td>Images</td>
<td>3.77 (4.13)</td>
</tr>
<tr>
<td>Block Screener</td>
<td>4.51 (4.13)</td>
</tr>
</tbody>
</table>

No statistically significant differences were found in food consumption levels between the MTurk-processed images and the Screener across all food categories.

Discussion

Principal Findings

This pilot study demonstrates the feasibility of collecting data on food intake through participant-captured and crowdsourced images. The method of photographing eating events in the context of an EMA study was generally
well-accepted and executed by the study participants. Most remembered to take pictures of their meals and snacks, and few reported difficulties in uploading the images. Importantly, uploading the photographs was not more difficult for participants than entering the responses. Images could be processed in a timely manner, and there was high agreement in the MTurk Worker count responses, particularly for fruits, salty snacks, and sweet snacks images, thereby supporting the feasibility of using MTurk for image classification. In addition, the weekly consumption estimated by the food images and the Screener was comparable.

Overall, vegetables were reportedly consumed more frequently compared with the other 3 food groups, and this was consistent between the 2 methods. Neither method aimed to measure the serving size but rather the frequency that these food groups were eaten. Both the MTurk instruction and the Screener aimed to measure the types of food eaten; therefore, in 1 meal or snack, multiple vegetables might have been present and been the driving force behind higher numbers.

This study further supports the use of participant-captured images to assess eating events. Compared with traditional EMA methods such as completing surveys or dietary journals, this study had comparable compliance rates: participants remembered to take photographs 77% (3.85/5) of the time [21,34,35]. Previous studies have used participant-captured food images and found high acceptability and data quality, with participant compliance rates ranging from 30% to 60% [16-18,36]. These studies had used dietitians to assess the energy intake or macronutrient levels from the images; therefore, our study demonstrates a novel approach of coupling participant-captured food images with crowdsourced image labeling. This method can accurately assess eating events in an EMA study without a time-intensive dietician review.

On average, batches of images took 1 hour and 45 minutes to process. Batches could be run simultaneously, allowing many images to be labeled in the same time frame. Theoretically, if a user could start all batches at the same time, all 4056 HITs could be accomplished within the same 1 hour and 45 minutes. The most limiting factor for time efficiency was the user’s ability to prepare and publish the batches of images. Even so, using MTurk to crowdsourced image identification is more time-efficient compared with the participants or study team members labeling the images. This is especially significant for future studies that wish to scale up this model with more participants or expand food labeling outside of the 4 food options examined in this study.

Agreement in Worker responses using MTurk was high for most food categories; however, for vegetables, they only had a moderate agreement. Among the images that received discordant vegetable responses, approximately 50% were because of image subject uncertainty, the highest across all food categories. In this discordance explanation category, the subject of the image presented a meal or snack that made it difficult to discern the count for the type of food present. Typically, this involved a mixture of foods with some items hidden beneath others or a variety that was difficult to differentiate. Vegetable discordance may have been higher as these foods are more often eaten mixed within foods, such as salads, guacamole, rice bowls, and soups, compared with the other food groups. Future work is needed to optimize the accuracy of vegetable intake using MTurk.

Crowdsourcing with MTurk has been used successfully in several areas of biomedical research, including endoscopic video image annotation [32] and optic disc image classification [31]. Other studies have reported the feasibility of using MTurk for crowdsourcing nutritional analysis from images of meals [30]. This system, called PlateMate, estimated the macronutrient calories from fat, carbohydrates, and proteins shown, which is comparable with assessments among trained dietitians. Compared with end-of-day recall, crowdsourcing limits participant burden and self-reflection. Crowdsourcing allows many Workers to label images at the same time, cutting down the total time it would take participants or study team members to process the images. This action removes the obligation from study participants, lessening their burden. It also limits participant reactivity by reflecting on their eating choices, thus biasing the results when participants change their behaviors because of study procedures. Our pilot study supports the feasibility of using crowdsourcing to process images and offers the potential enhancement of EMA studies by using crowdsourcing to accurately capture eating events in real time.

Limitations

The small sample size and limited socioeconomic diversity in our sample require replication in study populations with lower levels of education and income. The widespread availability of smartphones across socioeconomic groups supports our findings, despite the recognized digital divide [37]. This study used only female participants; therefore, further studies may benefit from recruiting both men and women to determine the feasibility of using smartphones and crowdsourcing to assess eating behaviors. Another limitation involved the inability to capture the daily frequency that participants ate a certain food item from the Screener; thus, the measure may not be directly analogous to the information captured with the food images. However, with the exception of fruits, multiple questions within the food categories were used to capture consumption, so our measures may more closely reflect instances rather than days.

Along the same lines, we were only able to evaluate and compare the number of times a participant ate a particular food with the food images, as opposed to assessing the serving size or the amount of the meal that was actually consumed. Most EMA studies are designed to assess eating behaviors on a micro timescale; thus, future studies would benefit from incorporating portion size as well. Seto et al [38] demonstrated the feasibility of using voice-annotated videos of meals and snacks to accurately capture portion sizes; however, trained dietitians were involved rather than crowdsourcing. The benefits of documenting portion size versus the time and costs of collecting and processing these data require further consideration depending upon the study aims.

Conclusions

This pilot study demonstrates the feasibility of using participant-captured images categorized through a crowdsourcing platform to accurately depict eating events. This
approach offers a potentially time-efficient and cost-effective strategy for EMA studies of this type. It can provide a richer breadth of data that reduces recall and reactivity biases in EMA studies compared with previous methods, such as dietary surveys and journals. It also offers an alternative strategy that is less burdensome to participants than previous EMA methods, as it reduces the amount of recall required by the participant. Our findings support the use of food images as a way of facilitating the growing interest in measuring food group frequency and general eating behaviors in a consumer-friendly manner with minimal additional burden.

Acknowledgments
This work was supported by the National Institutes of Health (K01HL133531, UL1TR001422). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Data collection was completed before SZ’s employment at the National Institutes of Health.

Authors’ Contributions
KH, SZ, LVH, and KK contributed to the design of the study. KH, LG, and NM analyzed the data. KH and KK drafted the manuscript. All the authors provided critical feedback on the manuscript and approved the final version.

Conflicts of Interest
None declared.

References


Abbreviations

CHEERS: Chicago Healthy Eating Environments and Resources Study  
EMA: ecological momentary assessment  
HIT: human intelligence task  
MTurk: Mechanical Turk

©Katharine Harrington, Shannon N Zenk, Linda Van Horn, Lauren Giurini, Nithya Mahakala, Kiarri N Kershaw. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.12.2021. 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.
Users’ Perceptions Toward mHealth Technologies for Health and Well-being Monitoring in Pregnancy Care: Qualitative Interview Study

Jane Li1*, PhD; David Silvera-Tawil1*, PhD; Marlien Varnfield2*, PhD; M Sazzad Hussain1*, PhD; Vanitha Math3*, MD

1Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation, Marsfield, Australia
2Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation, Herston, Australia
3Department of Obstetrics and Gynaecology, Gold Coast University Hospital, Gold Coast, Australia
*all authors contributed equally

Abstract

Background: Mobile health (mHealth) technologies, such as wearable sensors, smart health devices, and mobile apps, that are capable of supporting pregnancy care are emerging. Although mHealth could be used to facilitate the tracking of health changes during pregnancy, challenges remain in data collection compliance and technology engagement among pregnant women. Understanding the interests, preferences, and requirements of pregnant women and those of clinicians is needed when designing and introducing mHealth solutions for supporting pregnant women’s monitoring of health and risk factors throughout their pregnancy journey.

Objective: This study aims to understand clinicians’ and pregnant women’s perceptions on the potential use of mHealth, including factors that may influence their engagement with mHealth technologies and the implications for technology design and implementation.

Methods: A qualitative study using semistructured interviews was conducted with 4 pregnant women, 4 postnatal women, and 13 clinicians working in perinatal care.

Results: Clinicians perceived the potential benefit of mHealth in supporting different levels of health and well-being monitoring, risk assessment, and care provision in pregnancy care. Most pregnant and postnatal female participants were open to the use of wearables and health monitoring devices and were more likely to use these technologies if they knew that clinicians were monitoring their data. Although it was acknowledged that some pregnancy-related medical conditions are suitable for an mHealth model of remote monitoring, the clinical and technical challenges in the introduction of mHealth for pregnancy care were also identified. Incorporating appropriate health and well-being measures, intelligently detecting any abnormalities, and providing tailored information for pregnant women were the critical aspects, whereas usability and data privacy were among the main concerns of the participants. Moreover, this study highlighted the challenges of engaging pregnant women in longitudinal mHealth monitoring, the additional work required for clinicians to monitor the data, and the need for an evidence-based technical solution.

Conclusions: Clinical, technical, and practical factors associated with the use of mHealth to monitor health and well-being in pregnant women need to be considered during the design and feasibility evaluation stages. Technical solutions and appropriate strategies for motivating pregnant women are critical to supporting their long-term data collection compliance and engagement with mHealth technology during pregnancy.

(JMIR Form Res 2021;5(12):e28628) doi:10.2196/28628
KEYWORDS

pregnancy care; wearable sensors; mobile health; acceptance; mHealth service; design; mobile phone

Introduction

Background

Pregnancy is a normal physiological process, with most pregnancies progressing without any problems. However, pregnancy may pose many risks and complications (eg, gestational diabetes mellitus [GDM], preeclampsia, and mental health problems), which might greatly affect the health of the mother, fetus, or both [1,2]. Together with women’s existing medical conditions (eg, diabetes and hypertension), pregnancy-related complications can lead to adverse outcomes such as the loss of pregnancy by miscarriage, stillbirth, or low birth weight [1,2]. A healthy lifestyle is essential for the health of the mother and fetus and can potentially reduce the risk of maternal complications [3,4].

The pregnancy journey involves regular checkups that allow clinicians to monitor progress, identify potential risks, and provide general advice to encourage a healthy lifestyle [5]. Different factors may contribute to the likelihood of developing pregnancy-related conditions. Individual risk factors (eg, age and BMI), lifestyle patterns (eg, diet and physical activity), and physiological measures (eg, blood glucose levels, blood pressure, and proteinuria) are all indicators of pregnancy-related risk conditions [6,7]. Regular health and well-being monitoring can support early detection of health risks, improve treatment, and promote lifestyle adaptations in pregnant women [8,9].

Mobile health (mHealth), which involves the use of mobile and wireless technologies to support the achievement of health objectives [10], has been widely used in health care [11-13]. Wearable sensors and health monitoring devices are becoming popular and being used to support the monitoring of health and well-being [12-15]. These technologies support the sensing, tracking, and reporting of individuals’ health measures continuously (eg, physical activities and physiological data). Smartphone apps, coupled with wearable and sensing devices, have been used as data interfaces for visualization of measurement data, as motivational tools via persuasive messages, and to support personalized digital interventions to improve care programs [16,17]. With the availability of these technologies in health care, passive monitoring and personalized assessment would become integral to continuous patient monitoring [18].

The role of mHealth during pregnancy is being increasingly investigated [19-23]. Using apps to support pregnant women enhances the traditional pregnancy care model by providing additional educational information and empowering women to look after their own health [21,23-26]. Sensors and monitoring technologies that automatically track specific health indicators have been integrated into mHealth solutions to support pregnant women’s self-care behaviors [22,27-29].

Challenges in Pregnancy mHealth Care

Despite the expansion of mHealth, the practicality, design, and user needs for digital health monitoring in pregnancy require more attention [23,30]. There is a range of consumer-based wearable sensors and prototypes that can measure the physical activity, sleep, and physiological parameters [18,31,32]. However, none of these have been specifically designed for pregnancy care. While research has explored women’s and clinicians’ views of mHealth in pregnancy, there is still a knowledge gap regarding the preferences of mHealth monitoring among pregnant women and their clinicians as well as the suitability of mHealth monitoring for different conditions [24,29,30,33]. Recent studies have highlighted the importance of patient-centered design and behavior decision research in the development of mHealth solutions for pregnancy [19,23,28]. Understanding the women’s and clinicians’ preferences and their existing and preferred monitoring practices is crucial to assist in the design of practical solutions to promote a healthy lifestyle during pregnancy [23,24,28].

Intelligent data analysis can be used to identify the early signs of illness [34] and potentially support the early detection and management of complications in pregnancy [35,36]. Previous studies have investigated the use of predictive analytics and apps to support pregnancy care, with a focus on specific conditions and the medical data collected by clinicians [8,37]. These solutions required access to data in medical record systems and did not consider lifestyle (eg, sleep, diet, and exercise) factors in their analysis. The ability to extend the capability of clinical monitoring with multidimensional health and well-being data, collected via wearable and health monitoring devices, has the potential to provide significant benefits to the pregnancy journey. However, challenges remain in the collection of large-scale and long-term quality data deemed suitable for pregnancy care [19,23].

Related to these challenges is the need to support pregnant women’s engagement with technology [24,25,27,28,30]. Even with emerging evidence on the potential benefits, barriers to the implementation of mHealth technologies in pregnancy care persist [19,24,27,28,38]. Various factors can impact an individual’s feeling toward sharing and tracking health data, including stress associated with mHealth monitoring, the availability of reliable educational information, and ineffective communication with clinicians [22,25,27,30]. Further research is needed to understand pregnant women’s motivation to use technologies to better support their engagement, data collection compliance, and daily use [27,33]. This understanding can inform the development of mHealth lifestyle interventions and the integration of mHealth into pregnant women’s daily routine and the clinicians’ care practices [24,25,30,38].

Objective

As an initial step to inform the design of appropriate technologies to support the monitoring of health and well-being during pregnancy, we aim to conduct a qualitative study with clinicians, pregnant women, and postpartum women to understand their existing risk assessment and monitoring practices as well as their needs, interests, and preferences in mHealth. We also aim to explore the potential factors that may
influence their engagement with mHealth data collection and monitoring as part of pregnancy care. The study, conducted in the Australian context, contributes to the broader understanding of factors that motivate the use of mHealth in pregnancy care and how novel technologies can be designed and introduced to improve user engagement and long-term health monitoring during pregnancy.

**Methods**

**Overview**

A qualitative study was conducted using semistructured interviews with pregnant women, postpartum women, and clinicians currently working in pre- and postnatal care. The study was approved by the Commonwealth Scientific and Industrial Research Organization Health and Medical Human Research Ethics Committee (reference number: 2019_017_HREC) and the Gold Coast Hospital and Health Service Human Research and Ethics Committee (reference number: LNR/2019/QGC/54173) in Australia.

The interview questions were adapted from previous studies in conducting qualitative studies on digital health technology design [39,40]. The interviews with pregnant women and postpartum women included questions related to their experience in monitoring their health and well-being during their pregnancy, in particular their previous and current use of digital health technology (activity trackers and health monitoring devices and apps), and their interest and intention to use technology to support the management of their pregnancy needs. The motivations and factors that would contribute to their potential acceptance of mHealth monitoring were also explored. In addition, basic demographic and general information about their lifestyle during pregnancy was collected.

The clinician interviews included questions related to pregnancy risk assessment and management as well as supporting pregnancy health and well-being in standard clinical care. Clinicians’ thoughts were also collected on monitoring the components of mHealth solutions, on potential medical conditions for mHealth monitoring, and on how to introduce mHealth monitoring in pregnancy care. Basic demographic information about their roles and years of experience was also collected.

**Participants**

**Female Participants**

The criteria for pregnant and postpartum female participants (from here on, termed female participants) included the following: an age of ≥18 years, pregnancy (any stage of gestation) or postpartum pregnancy (no longer than 6 months postbirth) at the time of study, and the ability to give informed consent. The female participants were initially recruited via 2 internal email mailing lists within the authors’ organization. The email recipients were asked to share the invitation with friends and relatives who might be interested in participating. A snowballing technique was also used, asking women who had been interviewed to suggest other potential participants. Once an expression of interest was received, an information sheet and a consent form were sent to the potential participants via email, and interviews were scheduled after the participants consented. Purposive sampling was used during recruitment, with participants selected by considering their pregnancy stages. The focus was on a typical pregnancy, and high-risk pregnancies were not targeted. Initial data coding was performed during the data collection process. Recruitment continued until no new codes were identifiable in the subsequent interviews. The female participants were offered a gift voucher value of Aus $60 (US $42) as compensation for their time.

**Clinician Participants**

Clinicians recruited for the interviews were health care professionals involved in pre- and postnatal care at the obstetrics and gynecology department of a tertiary public hospital in Australia, which delivers standard clinical services in pre- and postnatal care. The management team of the department was interested in exploring the use of mHealth technology. Potential participants were identified by a key study representative (an obstetrician) at the hospital. They were chosen based on their role, level of experience, and interest in participation. The study representative emailed potential participants the information sheet and consent form to introduce the study and study investigators. Clinicians were required to express their interest in participating before being included in an interview. No compensation was offered to the clinician participants, except for a chocolate snack after the interview.

**Study Procedure**

All interviews were conducted on a one-on-one basis with each female participant during a 3-month study period. They were all offered the choices of face-to-face, videoconference, and telephone interviews. To ensure that the clinician participants were less inconvenienced, 2 researchers made themselves available at the hospital for 2 full days. Clinician participants could attend the interviews anytime during those 2 days, if their workloads allowed. Face-to-face interviews were conducted either at a clinician’s private office or at a small meeting room in the hospital. In addition, a telephone interview was offered to a clinician who visited a different hospital. The interviews were scheduled with the help of the study representatives. Each interview session was conducted with 1 clinician, except for 1 session involving 2 nurse managers. Each interview was conducted by 2 researchers and lasted approximately 30-45 minutes. The interviews were audio-recorded and transcribed professionally.

**Data Analysis**

A thematic coding technique was used to identify the insights from the interview data. One key researcher worked on the coding of the transcripts using NVivo (QSR International) data analysis software, and the other 2 researchers summarized their interview notes. Initial findings regarding the themes that emerged were discussed among the research team members. A second round of coding and analysis was conducted by a key researcher. The discussion continued over several meetings before a report of the findings was generated. Themes related to the current practices of regular monitoring and risk assessment, current experiences, and motivations for using...
mHealth technologies, and perceived benefits and challenges in incorporating mHealth technologies in daily life and practices were identified.

**Results**

**Participant Characteristics**

A total of 8 female participants were interviewed (Table 1); 2 (25%) of them were in the second trimester of pregnancy, 2 (25%) were in the third trimester, and 4 (50%) were in the postpartum stage. Pregnant women in their first trimester of pregnancy were not available during recruitment. Of the 8 female participants, 2 (25%) had GDM, 1 (13%) had preeclampsia, 1 (13%) had iron deficiency, and 4 (50%) had no medical conditions during their pregnancies. Moreover, 4 (50%) of the 8 female participants used private hospitals, 3 (38%) used public hospitals, and 1 (13%) had both public and private hospital experiences in their current and previous pregnancies. A total of 4 (50%) female participants were recruited from the researchers’ organization. All (8/8, 100%) of the female participants had experience using smartphones and mobile apps in general. A total of 13 clinicians from the public hospital were interviewed, including 2 (15%) obstetricians who also worked at private hospitals, 7 (54%) midwives in different roles, 1 (8%) health educator, 1 (8%) social worker, 1 (8%) physiotherapist, and 1 (8%) dietitian (Table 2).

<table>
<thead>
<tr>
<th>Table 1. Characteristics of female participants (N=8).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Characteristics</strong></td>
</tr>
<tr>
<td>Pregnancy stage</td>
</tr>
<tr>
<td>Second trimester</td>
</tr>
<tr>
<td>Third trimester</td>
</tr>
<tr>
<td>Postpartum (2-4 months)</td>
</tr>
<tr>
<td>Medical condition during pregnancy</td>
</tr>
<tr>
<td>Gestational diabetes</td>
</tr>
<tr>
<td>Preeclampsia</td>
</tr>
<tr>
<td>Iron deficiency</td>
</tr>
<tr>
<td>No medical conditions</td>
</tr>
<tr>
<td>Working time during pregnancy</td>
</tr>
<tr>
<td>Full-time</td>
</tr>
<tr>
<td>Part-time</td>
</tr>
<tr>
<td>First time mother</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>Public or private service used</td>
</tr>
<tr>
<td>Public</td>
</tr>
<tr>
<td>Private</td>
</tr>
<tr>
<td>Mixed</td>
</tr>
<tr>
<td>Have used smartphones and apps</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
</tr>
</tbody>
</table>
Table 2. Characteristics of clinician participants (N=13).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Roles</strong></td>
<td></td>
</tr>
<tr>
<td>Obstetricians</td>
<td>2 (15)</td>
</tr>
<tr>
<td>Midwife</td>
<td>3 (23)</td>
</tr>
<tr>
<td>Midwife, manager</td>
<td>2 (15)</td>
</tr>
<tr>
<td>Midwife, clinical consultant</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Midwife, general practitioner liaison</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Health educator</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Social worker</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Physiotherapist</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Dietitian</td>
<td>1 (8)</td>
</tr>
<tr>
<td><strong>Experience in their fields (years)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;10</td>
<td>2 (15)</td>
</tr>
<tr>
<td>10-20</td>
<td>6 (46)</td>
</tr>
<tr>
<td>20-30</td>
<td>4 (31)</td>
</tr>
<tr>
<td>&gt;30</td>
<td>1 (8)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>11 (85)</td>
</tr>
<tr>
<td>Male</td>
<td>2 (15)</td>
</tr>
</tbody>
</table>

In this section, we have presented the key themes extracted from the data, which are grouped into the following categories: risk assessment, health monitoring practices, and care needs; female participants’ experience of health monitoring and attitude toward mHealth; and the clinician participants’ perception of mHealth.

**Risk Assessment and Monitoring During Pregnancy**

Clinicians pointed out that every pregnant woman has unique care needs and potential risks. They articulated when and how risk assessment and support of lifestyle adaptions could be carried out in practice.

**Risk Assessment and Care Needs**

The need for longitudinal monitoring was highlighted by clinicians because many women who were initially believed to have low-risk pregnancies due to their medical history could become high-risk as their pregnancies progressed; one of the clinician participants (C9) commented “pregnancy is a journey with unpredictability.” As the clinicians explained, specific indications and risks for medical conditions were assessed during the early stages of pregnancy, throughout the pregnancy period, during labor, and during postpartum. In the early stages of pregnancy, the women’s health is typically monitored by their primary care provider, the general practitioner (GP), before their first visit to a hospital or an obstetrician. On the initial visit to a hospital (for women using the public system), medical data from the GP referral, medical history (eg, previous obstetric, family, and psychosocial histories), pre-existing disorders, and lifestyle (eg, smoking and alcohol intake) were consolidated. For women who used private hospitals, obstetricians managed their care and monitoring from the early stages of pregnancy until postpartum. In public hospitals, midwives used the guidelines of referral and consultation to decide how and when to refer women at different stages of pregnancy to obstetricians and other health care professionals. The care plan and monitoring arrangements depended on the women’s condition, their risk factors, their preferences, and the hospital’s capability. One obstetrician explained:

If anything changes along that timescale, they get referred back to obstetric clinic and they come under obstetric care, but that is because they developed a complication along the lines, either would be gestational diabetes, or they have problems with blood pressure or baby is not growing as well as it should be. [C11]

Assessment and care coordination is required for high-risk pregnancies. At the public hospital where the study was conducted, a clinical midwifery consultant worked as a navigator to provide coordination and consultation in collaboration with the obstetricians and other specialists for high-risk pregnant women. Such cases were flagged in the electronic medical records and their management plans were recorded. The high-risk pregnant women need to visit the hospital multiple times to see different specialists, whereas others (intermediate risk) whose conditions are well-managed only need some level of coordination. The changing nature and different care needs for pregnant women with different risk profiles are described as follows:

It really depends on the health and well-being of the women, and sometimes I can downgrade women and sometimes I have to upgrade them as their pregnancy goes along. So sometimes, women at the beginning
need quite a lot of intensive support and coordination, but once we have got that underway and we are on the right track, then it can be stepped down. [C13]

**Supporting Pregnant Women’s Healthy Lifestyle and Well-being**

Clinicians pointed out that awareness and support of physical activity, diet, and mental and emotional aspects of well-being in pregnant women is an important component of pregnancy care. Nutrition, exercise needs, and expected weight gain were discussed with the pregnant women during their first visit. This information was included in the information pack, together with useful website information and web links. Although the low-risk women received general education, specialist consultations were provided for at-risk women (eg, those with diabetes) to guide their diet and physical exercise. Physiotherapy and dietitian support were provided for women who needed interventions. Mental health was assessed by an obstetrician review or a midwife review. The Edinburgh questionnaire [41] was used as a screening tool. Triage assessment for women with high scores was performed, followed by a referral to the mental health team for their ongoing management during the antenatal and postnatal periods.

The midwife manager participants pointed out that there has been some attention to monitor women’s lifestyle behavior, emotional health, and well-being from a longitudinal perspective. There is a trend of expanding the timeline of pregnancy and perinatal care to the first 1000 days from conception to when children turn 2 years old [42]. This is a critical period as the pregnant and postnatal mothers’ health, nutrition, and stress levels can have a long-term impact:

> The education around that for the mother...is modifiable behaviours...giving them the right support and the right education that can make a big difference, even the small changes...for the length of stay, and for the long-term health benefits of that child if they can change those behaviours when they go home as well. [C7]

**Female Participants’ Perception on Health Monitoring and mHealth**

**Overview**

All female participants believed that maintaining a healthy lifestyle was beneficial for their pregnancy. Most of them integrated a certain level of exercise into their daily routines. Walking was the most popular type of activity among them, followed by activities such as swimming and aerobics. They also maintained a balanced diet and monitored their body weight.

In this section, we have described the female participants’ views on health measurements (eg, blood pressure and blood glucose levels) and lifestyle behaviors (eg, physical activity, diet, sleep, stress and mental health, and weight management) that can be tracked using wearable sensor devices, health monitoring devices, and mobile apps.

**Monitoring Physical Activity, Sleep, and Heart Rate**

Commercial activity trackers (eg, Fitbit [Fitbit Inc] and Apple Watch [Apple Inc]) were used by 63% (5/8) of the female participants to track their steps, sleep, and heart rate. Of the 5 of them, 2 (40%) used activity trackers before pregnancy and continued using them during pregnancy, whereas the other 3 (60%) bought a device specifically to monitor their physical activity during pregnancy. Participants who did not own or use activity trackers during pregnancy thought that they had no medical problems, were physically active already and did not need additional motivation, or were concerned about the inconvenience of wearing a tracker and the need to charge the device battery.

Physical activity was the key measure tracked by the participants who used trackers. Furthermore, 25% (2/8) of participants used trackers to keep track of sleep quality. Interestingly, although 4 (50%) of our 8 participants mentioned sleep problems during pregnancy (eg, waking up a couple of times at night and difficulty going back to sleep), they did not feel the need to track their sleep every day as they knew they had this issue. Similarly, heart rate was not a concern for most of them, with only 25% (2/8) of participants who tracked their heart rates on the activity trackers noticing an increased heart rate when they were stressed:

> I tended to notice that it seemed to be when I was trying to rush around somewhere or I was a little bit anxious, it was often higher when I was at the hospital...so I was just conscious to kind of take some deep breaths and just try and sit down and calm down for a little bit. [F2]

**Monitoring Weight**

All female participants used weight scales at home before and during their pregnancy. Of the 8 participants, 2 (25%) paid particular attention to weight increase as suggested by doctors, whereas another 6 (75%) participants used the scales on a now and then basis:

> I didn’t properly track my weight, I just put myself on the scale every week or so...And the obstetrician always got my weight as well on his scales, so he tracked it that way. [F5]

**Monitoring Diet**

Participants with GDM diagnosis tracked their calories. Of the 8 participants, 1 (25%) tracked food intake using an app and found it useful, as she was told by her physician not to gain more than a certain weight. Half (4/8, 50%) of the participants expressed their willingness to try a diet-tracking app. The other half felt that they did not have any major health concerns that required them to track food intake, that their weight was in healthy range, or that they were concerned about the time and effort required to record and check the data:

> That's never been a priority for me to monitor how many calories I’ve had during the day because I’ve never been someone that over eats and I’ve always stayed within a pretty healthy weight range. [F3]
Monitoring Blood Pressure

Most of the female participants did not measure their blood pressure at home, with it being measured only during their hospital visits. Only 38% (3/8) of them had a blood pressure device at home, with only 25% (1/4) of them (who used a private hospital) measuring it regularly at home owing to abnormal blood pressure being detected occasionally on her visits to the clinic. Participants were educated by their clinicians on how to monitor their symptoms and were informed that their local GPs or pharmacists could check their blood pressure if needed:

If I had other symptoms, I felt confident that I could just find that out quite easily and quickly, and I know a lot of pharmacies and things they can just do a quick blood pressure check. [F1]

Monitoring Blood Glucose Level

Female participants with GDM maintained regular self-measurement and reporting of blood glucose levels during pregnancy. The readings from a blood glucose testing device were recorded by women (in a public hospital) in a booklet and discussed with clinicians on their visits. At private hospitals, this was recorded by women on an Excel (Microsoft, Inc) spreadsheet and emailed to the clinics every 1 or 2 days to be reviewed by the hospital’s endocrine or obstetrics team.

Monitoring Mental Health

The difficulty in receiving mental health support during the first trimester was expressed by 63% (5/8) of participants, as they were reluctant to talk about their pregnancy at this stage. Some of them had morning sickness and did not enjoy food or exercise. Although a GP is the primary care provider before their first appointment at the hospital (or obstetric clinic), most women had not established regular GP visits in their first trimester.

In terms of tracking mood, it appeared that 25% (2/8) of female participants who had good support network (family and friends) were less interested in tracking their mood, whereas most participants expressed this need as they either felt stressed in maintaining their work commitments or became very emotional during pregnancy:

Problems that probably didn’t seem like a problem before all of a sudden seem so much worse, that sort of gets you down a bit –that sort of feeling like you emotional and you really want to talk about something and – but then you wake up in the morning and you’re like I don’t know what I was upset about, it’s probably that sort of feeling that I’ve had throughout the pregnancy. [F3]

Using Mobile Apps

All female participants had experience using various apps that provided pregnancy and postnatal information or assisted in the tracking of fetal movements. The key function they used was to obtain information, such as week-by-week information about their pregnancy progress and the baby’s growth, the symptoms to look out for, nutrition information, and mental health support. Apps for postnatal care were also used by 25% (2/8) of participants to track the feeding, sleep, and growth of their babies. Most participants used the apps to receive information than to enter information:

I actually don’t put much information into it, I use it more just for information like sourcing, but it does allow you to track all your appointments and put all your symptoms in and things like that as well...But just having to enter information every day without kind of getting any information back I don’t know that wouldn’t be so appealing. [F1]

Female Participants’ Attitude Toward mHealth

Female participants’ overall attitude and concerns toward mHealth included the following:

- They would be motivated to use mHealth if they knew that clinicians could access their mHealth data, and they were willing to share and discuss the data with clinicians at the hospital or clinic visits.
- They would prefer to use devices that featured automatic data capture without the need for manual data entry.
- If the use of monitoring devices (eg, blood pressure and blood glucose level) had the potential to be associated with positive outcomes, they would have had a stronger motivation to use them.
- They would be less motivated to use mHealth if they had no medical conditions or potential risks during pregnancy.
- They were concerned that it could be a source of anxiety if the measurements were slightly out of the normal range, adding to the stress they already had during their pregnancy.
- Half of them believed that most of the apps managed data privacy and security well, whereas the other half were concerned about potential security issues with their data. All of them indicated that if there is an assurance of data security and proper use of the data (such as studies aimed at improving pregnant women’s health), they would be more likely to track lifestyle behaviors and health measures.

Clinician Participants’ Perception on mHealth

Overview

Clinician participants saw the potential of using mHealth technologies for the monitoring of health and well-being in pregnant women, particularly for longitudinal monitoring during pregnancy, as complications could develop in women with or without a risk history. They also acknowledged that it could play an important role in supporting the current practices of risk assessment and care for pregnant women with different risk levels. According to them, for pregnant women classified as high-risk and for those who required additional education and monitoring of their health status, mHealth can be an invaluable tool to improve their compliance.

In this section, we have described the clinicians’ views on technology requirements and suitable conditions for mHealth monitoring and the factors that may help introduce technology in their practice.

Broad Requirements

Data collection from multiple sources, incorporating accurate information to pregnant women, clinician portal to access data,
providing features for alertness, and ease of use were among the desired features for the clinicians.

**Health Monitoring Data**

Clinicians agreed that some generic health parameters, such as blood glucose and weight, could be measured by pregnant women at home. The dietitian participant mentioned that a wearable continuous glucose monitoring device was provided to patients with type 1 diabetes and this was supported by the Australian government’s continuous glucose monitoring initiative [43].

The clinicians’ opinions on the measurement of blood pressure at home by pregnant women varied. Of the 13 female participants, 3 (23%) of them were supportive, whereas others expressed concerns. In total, 6 participants explained that misinterpretation and inaccurate readings can cause unnecessary anxiety as pregnant women are not trained to measure blood pressure correctly; individuals’ reference ranges might also be slightly different for each woman, and clinicians did not expect low-risk women to take their blood pressure at all times. One clinician participant commented the following:

> Because they use electronic blood pressure cuffs at home, they are not trained to use a manual one and the electronic ones may not accurate, they can be a bit off and then that could lead to the woman starting to worry and think—oh my blood pressure is so high... If needed we might get them to go to their GP and get it done more frequently... or monitoring blood pressure by visiting the pharmacy. [C8]

**Lifestyle Data**

Clinicians were most receptive to the use of physical activities, diet, and sleep monitoring when considering wearable sensors and apps for pregnant women. One clinician mentioned that some existing blood glucose monitoring apps can collect additional information such as self-reported insulin dose, dietary intake, carbohydrate amounts consumed, and the time of exercise. Another clinician pointed to the link between good sleep and clinical outcomes, such as blood glucose level control. Physical activity tracking was considered particularly useful for women with a high BMI or with diabetes:

> Physical activity is helpful for women, but in terms of us it is not something we ask normal women like we won’t say have you been walking three times this week, but it is relevant if she has diabetes where she needs to do the physical exercise of if she has a higher BMI. [C8]

**Questionnaire Data**

Incorporating validated questionnaires (such as the Edinburgh questionnaire to measure the risk of mental health issues) into an mHealth solution was suggested by some clinicians. However, they also pointed out that the questionnaire results would need to be analyzed in combination with other measures. The frequency of the questionnaire should be considered on a case-by-case basis, and guidelines were needed for clinicians to follow up on the results, as one clinician explained:

> We have to have some process of being able to pick that up and work with, say if someone reports that they’re not doing so well or they have suicidal ideation we have to make sure we have clear pathways of what to do with that information. [C6]

**Incorporating Accurate Information and Feedback**

Clinicians saw the importance of incorporating an app providing tailored information and feedback to pregnant women into the mHealth technical solution.

According to the clinicians, pregnancy is a process of education and information seeking for pregnant women. Not all pregnant women read the material provided by hospitals. Some might not know their risks, the consequences of the risks, and the symptoms to watch. They might seek materials from the internet or educational content from the available apps that provide general information. Hospitals did not provide suggestions for the selection of apps.

Clinicians highlighted that educational content needs to be accurate and tailored to particular conditions. Ideally, it should provide individualized information or advice to women with different risk factors and should integrate targeted information as a component to encourage the women’s use of mHealth during pregnancy. One clinician said, “It’s just got to be continually meeting the needs of the different cohorts and the health literacy of the individual” (C7).

**Access Data by Clinicians**

Clinicians discussed the need for a portal or central source of mHealth data for different clinicians to access. The ability to review the data can better support patient-clinician communication and improve the efficiency of face-to-face consultations.

**Alerts**

The clinicians were supportive of an alert feature. They pointed out that one key aim of monitoring should be to make women respond to their data, that is, women getting flagged by an alert that could trigger their access to health professionals. Predictive analysis based on the monitoring data could also potentially provide alerts to clinicians to allow early detection of problems and timely interventions.

**Ease of Use**

Clinicians highlighted that mHealth technology needed to be user-friendly and with minimum effort to use for pregnant women. They suggested that mHealth solutions incorporate monitoring devices with the feature of automatic data capture.

**Clinical Considerations for mHealth Monitoring**

The conditions in pregnant women groups that can potentially benefit from mHealth technology-assisted monitoring were discussed by the clinicians.

**Gestational Diabetes**

Pregnant women diagnosed with GDM receive ongoing support from diabetes educators, diabetes dietitians, and endocrine specialists from the pregnancy to postnatal stages. According to our clinician participants, women with GDM were advised...
to test their glucose levels at home 4 times a day using a glucose monitor, record the results, and have periodic follow-up visits at the hospital. Existing glucose monitoring practices require extensive effort from both female patients and clinicians. Clinicians could benefit from easy access to data through web-based or mobile solutions. Our clinician participants indicated that GDM is the most common and suitable medical condition to consider for mHealth interventions.

**Hypertension**

For women with hypertension, blood pressure needs to be monitored for potential risks of preeclampsia, or to assess the patient’s response when taking medication for high blood pressure. It is important to provide training to these women on how to use blood pressure monitoring devices. Blood pressure measures also need to be assessed in combination with symptoms such as headache, blurred vision, and swelling as well as pathology tests.

**Obesity or High BMI**

Women with a high BMI might have pre-existing diabetes or are at a high risk for gestational diabetes. It is important for them to maintain an appropriate lifestyle. They can benefit from regular weight tracking and diet monitoring.

**Mental Health**

This includes antenatal and postnatal depression, anxiety, and depression. mHealth has the potential to help track how pregnant women feel, their mood, and when to receive timely intervention and counseling.

**Stillbirth Prevention**

Obstetrician participants pointed out that despite previous efforts to improve care and monitoring, progress in reducing stillbirth rates remains low. Sleep apnea can be a risk factor for stillbirth, and obstetricians were interested in investigating the relationship between sleep abnormalities and stillbirth using smart sensor technologies.

**Other**

Women with previous history of pregnancy problems, such as fetal growth restriction and fetal loss, were discussed by the clinicians.

**Challenges**

Although the clinicians were positive about the potential uses of mHealth technologies, challenges around engaging pregnant women, technology issues, changes in practices, and evidence-based solutions were discussed in the interviews.

**Women’s Engagement**

According to clinician participants, pregnant women’s compliance with self-reporting (such as diet and questionnaires) could be a challenge, especially for women who were busy and if the recording process was not simple. Women might stop using the monitoring devices if monitoring created anxiety. Language and cultural barriers can be an issue for the engagement of non–English-speaking groups. Although pregnant women diagnosed as high-risk tend to be more compliant, providing education and showing benefits can potentially improve data collection compliance for other women:

*I would worry about the compliance that was going to be my first feedback...their lives are so busy and their stress levels fluctuate a lot and they would find it hard to commit all the time...I think the compliancy is as good as the education they’re given, if we explain it well and how it can benefit them and empower them...So the biggest blockage for technology is compliance and consistency. [C3]*

**Technology**

Technology concerns were captured during the interviews. These included accuracy of wearable sensors and devices, as it was directly linked to the reliability of the data, and the cost of a device, as it was an issue for women with low income if the device was expensive. Therefore, providing individualized information was a challenge, as it would be difficult to meet the complex needs of different women and their different conditions.

Intelligent modeling for prediction and its accuracy can be challenging. From a medical perspective, predicting a medical condition is not easy, according to one obstetrician:

*At this stage, in terms of finding predictors, in the first trimester, second, early second trimester, that would predict things like preeclampsia, things like gestational diabetes, things like growth restriction or foetal demise, even now they're still not there, they don’t exist. [C11]*

An obstetrician pointed out that detection of clinical abnormalities can be enhanced by including other clinical and health data from electronic health record systems. However, the integration with other record systems can be a challenge.

**Change of Practices**

Few self-reported measures (except for blood glucose level for patients with diabetes and weight tracking at some hospitals) were collected in standard practice at the time of our study. The midwives did not collect objective physical activity data and diet information from low-risk women.

Reviewing the women’s monitoring data when introducing mHealth was raised as a concern by some clinician participants (eg, obstetricians) as it would require extra work in their already busy schedules. However, other clinicians (eg, midwives) responded that this would not be an issue for them, but it might need a dedicated staff member to take the responsibility and time to check the data and follow up when needed. Intelligent decision-making with alerts for abnormal measures was considered helpful for clinicians.

There is also a need to improve communication between clinicians from multiple disciplines, including GPs. Engaging busy clinicians by showing potential outcome improvements could motivate them to be involved:

*All clinicians want to do the right thing but only have a limited amount of time, so I think you would engage clinicians by showing them the data on...literature...and (explain) that if we do this, women are less likely to end up having this outcome...I think people would be excited for that. [C2]*
Demonstration of Impact

Clinician participants supported the approach of conducting an evidence-based trial before introducing mHealth into practice. Most clinicians were willing to participate in an mHealth trial. The initial steps they suggested included targeting particular groups (eg, women in rural and remote areas) and particular conditions and supporting women from low socioeconomic backgrounds where the prevalence of risk factors is common, and technology can make a difference:

So I think in the right conditions with the right people it would work. I could see that it would certainly work in some of the rural and remote areas...if they could just with an App send in their information and then someone can look at it and just ring them up and reassure them. [C9]

Discussion

Potential Interests

mHealth technologies for health and lifestyle monitoring have been used in the general population. There is a growing interest in introducing mHealth solutions to support the pregnancy journey, which is the period of a woman’s life that involves significant physiological changes and potential risks [29,44]. In this study, we examined the interests and perceptions of women and clinicians regarding the use of mHealth for health and well-being monitoring during pregnancy.

Our study showed that female participants were open to the use of wearables and health monitoring devices to track health and well-being in general, with most of them having previous experience of using physical activity trackers and mobile apps before and during their pregnancy. This result echoes previous research [25,30], including a study conducted in the Australian context [24]. Despite low interest in monitoring lifestyle behaviors among low-risk pregnant women with no medical problems, all female participants felt comfortable sharing information from wearable and monitoring devices with their clinicians and would have felt motivated if clinicians could review the data. In addition, women with a GDM diagnosis were normally engaged in continuous health monitoring of blood glucose, with data being recorded manually, and would be supportive of an mHealth solution to make the process more efficient.

Clinicians in our study did not use mHealth technology and wearables or prescribe mobile apps in the current practices. Our findings revealed that there was an overall positive response among these clinicians on the potential benefit of mHealth for monitoring pregnant women’s health and well-being and promoting healthy lifestyle behaviors during pregnancy, similar to the findings of other studies [24]. Despite the concern about women’s anxiety caused by self-interpretation of the data, clinicians were interested in using mHealth monitoring to assist the current practices of risk assessment and regular checks with pregnant women. Our study also highlights that mHealth monitoring is aligned with the trend of extending the context of perinatal care to a longitudinal health and well-being care model [26,42]. It can further serve to empower pregnant women to take more of an active role in their lifestyle behaviors during the antenatal, pregnancy, and postnatal periods.

Improving Engagement With mHealth Technology

Overview

Building a rich and multidimensional data pool is required to identify changes in lifestyle, health indicators, and risk factors associated with pregnancy complications. However, the motivation and sustainability of long-term data collection in pregnant women might be difficult [19,27,30]. We found that different contexts (eg, health status and access to support network) can impact an individual’s decision to track and share data. Our findings have shown that higher compliance can be achieved in women who were already engaged with their care (eg, women with higher risk) and women who embraced technology. Women with busy work commitments were less likely to comply with the use of mHealth solutions.

Design Considerations

Our study has shown that monitoring requirements and care needs vary with the combination of particular conditions and risk levels among pregnant women. As such, technology solutions need to be tailored to the unique needs as per the conditions and risk levels of the individual women. Different modules with different monitoring parameters and monitoring frequencies can be made available for clinicians to select and assign to women based on the severity of their conditions and risks. For women considered to be at high risk or for those with an available diagnosis, the focus of the solution can be on using condition-specific devices and parameters to help prevent adverse events and provide alerts to both pregnant women and their clinicians. For low-risk women, to reduce their unnecessary burden and anxiety, the focus of the solution should be to help them establish healthy lifestyle behaviors and watch their symptoms without the daily collection of medical data.

Irrespective of the risk levels, our research suggests that women will benefit from a mobile app that not only interfaces with monitoring devices but also provides guidance on healthy lifestyle and behavior changes. Other studies have shown that mHealth interventions often require support from other modalities, such as educational content [45,46]. Our study has revealed further details about the women’s tendency to seek trustworthy tools that deliver answers to weekly pregnancy and baby growth information, concerns in early pregnancy stage, information support services, and personalized information, such as nutrition, fitness, and weight. Pregnant women require clinically accurate and actionable information and feedback. Simple, engaging, tailored, and risk-appropriate information and text messaging delivered according to their stage of pregnancy can be useful in maintaining pregnant women’s interests and satisfaction. Similarly, motivation tools such as medals and rewards in apps can provide them with encouragement for achievements, such as targeted physical exercises or healthier gestational weight gain.

Implementation Considerations

Our study suggests some strategies on how to work best with less motivated pregnant women. First, one possible solution to...
help overcome this challenge would be to introduce the technology to women during the first trimester, which is a difficult stage when mental health support and self-guided information seeking are needed. This would allow them ample time to get comfortable with the technology and overcome some level of anxiety, thus motivating them for continued use in the later stage. Second, clinicians’ recommendations and indications of potential positive outcomes in women can help improve their acceptance. This may contribute to lesser anxiety and stress and higher motivation and reassurance for the women if they know that clinicians are involved. Finally, providing education and training to women in using technology is also important to reduce unnecessary stress and anxiety associated with mHealth during pregnancy.

Introducing mHealth Technology in Practices

Overview

Given the complexity of pregnancy care, there are challenges in introducing mHealth monitoring in care practices. Detecting clinical abnormalities and analysis based on high volume and heterogeneous data generated from mHealth devices can be challenging. This requires the skills necessary to accurately analyze the data for sound clinical decision-making. Participants of this study were also concerned about the extra workload for clinicians in data monitoring.

Design Considerations

Female participants and clinician participants of this study were supportive of having an mHealth system with an alerting function that could not only notify the clinicians of changes in a woman’s condition but also enable the women to be aware of problems and to be proactive in seeking professional service. Research in advancing data mining techniques and personalized algorithms has made intelligent detection and risk awareness possible.

However, according to our interviewed clinicians, accurately predicting the likelihood of a pregnancy risk and change in a condition is difficult in pregnancy care. It might require a multidisciplinary approach that considers pregnancy risk factors, symptoms, laboratory findings, and even data about the baby.

Health monitoring using physiological and activity measures from wearable sensors has been growing recently, but the integration of these technologies into practices, particularly pregnancy care, has been limited due to concerns about patient privacy, uncertainty about the reliability of the technologies, and usability, as reported in other studies [29,30,33,38]. In this study, the women’s views on privacy varied. Some women were not worried about it, whereas others were cautious about providing their data because of concerns regarding the maintenance of confidentiality for the captured data. Uncertainty in the reliability of these emerging wearables was also expressed by the clinician participants. Ease of use and automatic data capture were among the women’s and clinicians’ requirements for the devices. Technology development in truly wearable, miniaturized, and nonintrusive technologies can lower the barrier of usability and allow passive and longitudinal data collection.

Implementation Considerations

In this study, the clinician participants anticipated that some medical conditions such as GDM, hypertension, and mental health could benefit from the use of mHealth monitoring. However, evidence on the effectiveness of mHealth monitoring in pregnancy is limited and needs further investigation before supporting its future use. They suggested that some medical conditions (eg, GDM) and groups (women in rural and remote areas) would be suitable for the feasibility trials of mHealth and for further investigations before implementation. Longitudinal studies are needed to evaluate the efficacy of mHealth solutions for monitoring during pregnancy, especially in high-risk pregnancies as well as acceptability among pregnant women and clinicians to promote the uptake of mHealth technology.

Limitations

Due to the constraints in the hospital ethics application process for studies involving pregnant patients in hospitals, we were not able to recruit female participants from the hospital during the study period. All female participants were recruited through community advertisements and word of mouth. As such, the number of female participants was limited, particularly those in the first trimester of their pregnancy. Additionally, the clinician participants were recruited from a public hospital, although the obstetricians also worked at private hospitals. To enrich the current findings, further studies could gain insights from more clinicians working in private hospitals and the GPs. Finally, in this study, we only captured limited socioeconomic information from female participants. We found that the participants touched upon (only slightly) the challenges for women with low income or women with diverse cultural and linguistic backgrounds during the interviews. Future research on the impact of pregnant women’s socioeconomic status and cultural background might be needed to better understand the technology generalizability and digital equity in mHealth for pregnancy care.

Conclusions

We have explored the aspects of current risk assessment practices, users’ motivations, and concerns as well as clinical and technical factors that need to be considered when designing and introducing mHealth monitoring solutions for pregnant care. Adequate high-quality data collected through longitudinal monitoring is required for the intelligent detection of risks. We discussed technology solutions and implementation strategies to improve pregnant women’s engagement with technology and data collection, which are critical for mHealth solutions to facilitate the tracking of health and behavior changes during pregnancy. Future research will include feasibility studies to inform the development of mHealth technology and evidence-based evaluation studies to understand the efficacy of mHealth solutions in supporting pregnancy care.
Acknowledgments
This project was funded by the Commonwealth Scientific and Industrial Research Organization Precision Health Future Science Platform. The authors thank the clinicians at the Gold Coast University Hospital for their support in this study.

Conflicts of Interest
None declared.

References
21. Overdijkink SB, Velu AV, Rosman AN, van Beukering MD, Kok M, Steegers-Theunissen RP. The usability and effectiveness of mobile health technology-based lifestyle and medical intervention apps supporting health care during pregnancy:

https://formative.jmir.org/2021/12/e28628

JMIR Form Res 2021 | vol. 5 | iss. 12 | e28628 | p.100

(page number not for citation purposes)
systematic review. JMIR Mhealth Uhealth 2018 Apr 24;6(4):e109 [FREE Full text] [doi: 10.2196/mhealth.8834] [Medline: 29691216]


35. Li et alJMIR FORMATIVE RESEARCH


Abbreviations

GDM: gestational diabetes mellitus
GP: general practitioner
mHealth: mobile health

©Jane Li, David Silvera-Tawil, Marlien Varnfield, M Sazzad Hussain, Vanitha Math. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.12.2021. 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.
Initial Feasibility of the “Families Moving Forward Connect” Mobile Health Intervention for Caregivers of Children With Fetal Alcohol Spectrum Disorders: Mixed Method Evaluation Within a Systematic User-Centered Design Approach

Christie Lynn McGee Petrenko¹, BA, BSc, MSc, PhD; Carson Christine Kautz-Turnbull¹, BA; Alicia Rose Roth¹, BA; Jennifer Elizabeth Parr¹, BA, MSc; Cristiano Tapparello², BSc, MSc, PhD; Utku Demir², BSc, MSc, PhD; Heather Carmichael Olson¹, BA, MA, PhD

¹Mt. Hope Family Center, University of Rochester, Rochester, NY, United States
²Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, United States
³Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, WA, United States
⁴Center for Child Health, Behavior, and Development, Seattle Children’s Research Institute, Seattle, WA, United States

Corresponding Author:
Christie Lynn McGee Petrenko, BA, BSc, MSc, PhD
Mt. Hope Family Center
University of Rochester
187 Edinburgh St
Rochester, NY, 14608
United States
Phone: 1 585 275 2991
Email: christie.petrenko@rochester.edu

Abstract

Background: Fetal alcohol spectrum disorders (FASD) are prevalent neurodevelopmental conditions. Significant barriers prevent family access to FASD-informed care. To improve accessibility, a scalable mobile health intervention for caregivers of children with FASD is under development. The app, called Families Moving Forward (FMF) Connect, is derived from the FMF Program, a parenting intervention tailored for FASD. FMF Connect has 5 components: Learning Modules, Family Forum, Library, Notebook, and Dashboard.

Objective: This study assesses the feasibility of FMF Connect intervention prototypes. This includes examining app usage data and evaluating user experience to guide further refinements.

Methods: Two rounds of beta-testing were conducted as part of a systematic approach to the development and evaluation of FMF Connect: (1) an iOS prototype was tested with 20 caregivers of children (aged 3-17 years) with FASD and 17 providers for the first round (April-May 2019) and (2) iOS and Android prototypes were tested with 25 caregivers and 1 provider for the second round (November-December 2019). After each 6-week trial, focus groups or individual interviews were completed. Usage analytics and thematic analysis were used to address feasibility objectives.

Results: Across beta-test trials, 84% (38/45) of caregivers and 94% (17/18) of providers installed the FMF Connect app. Technological issues were tracked in real time with updates to address problems and expand app functionalities. On use days, caregivers averaged 20 minutes using the app; most of the time was spent watching videos in Learning Modules. Caregiver engagement with the Learning Modules varied across 5 usage pattern tiers. Overall, 67% (30/45) of caregivers posted at least once in the Family Forum. Interviews were completed by 26 caregivers and 16 providers. App evaluations generally did not differ according to usage pattern tier or demographic characteristics. Globally, app users were very positive, with 2.5 times more positive-than negative-coded segments across participants. Positive evaluations emphasized the benefits of accessible information and practical utility of the app. Informational and video content were described as especially valuable to caregivers. A number of affective and social benefits of the app were identified, aligning well with the caregivers’ stated motivators for app use. Negative evaluations of user experience generally emphasized technical and navigational aspects. Refinements were made on the basis of feedback during the first beta test, which were positively received during the second round. Participants offered many valuable recommendations for continuing app refinement, which is useful in improving user experience.
Conclusions: The results demonstrate that the FMF Connect intervention is acceptable and feasible for caregivers raising children with FASD. They will guide subsequent app refinement before large-scale randomized testing. This study used a systematic, user-centered design approach for app development and evaluation. The approach used here may illustrate a model that can broadly inform the development of mobile health and digital parenting interventions.

(JMIR Form Res 2021;5(12):e29687) doi:10.2196/29687

KEYWORDS
fetal alcohol spectrum disorders; fetal alcohol syndrome; intervention; mobile health; mHealth; parenting; children; prenatal alcohol; digital health; user-centered design; mobile phone

Introduction

Background

Fetal alcohol spectrum disorders (FASD) are a range of conditions associated with prenatal alcohol exposure (PAE) [1]. PAE can affect the development of the brain and other organ systems, resulting in neurodevelopmental impairment and high rates of physical and mental health problems [2,3]. FASD are highly prevalent, occurring in an estimated 1%-5% of the US population [4]. Unfortunately, FASD are markedly underrecognized, and most individuals do not receive an accurate diagnosis [4,5]. Families often only know an individual was exposed to PAE and is showing learning or behavior concerns. Many barriers to care exist because of the pervasive lack of knowledge about FASD across service systems and in the broader community [6,7].

People with FASD have important strengths, including social motivation, resilience, and individual passions and talents [8,9]. They strive to be included and contribute meaningfully to their communities. Caregivers (ie, biological, foster, adoptive, and relatives) are dedicated to supporting their children with FASD or PAE and undertake numerous protective actions to reduce system barriers and help their own children and families adapt to challenges [10-12]. However, responding to FASD remains a very stressful experience, often fraught with difficulties accessing the resources and information needed. A growing number of evidence-based interventions have been studied in preschool and school-aged children with FASD [13,14]. It is unfortunate that none of these are, as yet, widely available in community settings. Thus, innovative and scalable solutions are required.

A Systematic, User-Centered Design Approach to App Development and Evaluation: The Example of Families Moving Forward Connect Programmatic Research

To address significant barriers to care affecting this population, we developed a mobile health (mHealth) intervention called Families Moving Forward (FMF) Connect. mHealth, or the application of smartphones or wireless technologies to improve health, has burgeoned since the emergence of app stores in 2008 [15]. mHealth has many potential advantages, including increasing health care capacity, providing patient access to tailored and immediate support, reducing stigma in obtaining care, and improving cost-effectiveness [16]. FMF Connect is the first known mHealth app developed and tested for FASD. The task of developing and evaluating the FMF Connect mHealth intervention is being carried out following a systematic, user-centered design approach to app development and evaluation (Figure 1). Unfortunately, deployment of this type of systematic approach has been relatively rare for mHealth interventions [17-19]. This methodology integrates user-centered design principles, which emphasize understanding users, tasks, and environments, with the process of obtaining iterative and collaborative input from users [20-22]. There are seven phases to this approach, as operationalized in Figure 1. Of course, the process is more complex than that illustrated in Figure 1. There are certainly feedback loops between phases that indicate iterative change. This study describes the model, which involves a multidisciplinary development team and engagement of key stakeholders through focus groups and beta-testing. We also include the presentation of data from the fifth phase of the model to reveal specifically how user data from beta-testing can strategically refine and enhance app design. We note that this generalized approach can be used in the broader field of mHealth development.

In the first phase of this approach, the self-directed FMF Connect app was carefully derived from the empirically supported, therapist-led FMF Program developed by Heather Carmichael Olson, PhD and colleagues at the Seattle Children’s Research Institute and the University of Washington [23-25]. The standard FMF Program integrates clinical techniques of positive behavior support, cognitive behavioral techniques, and motivational interviewing to improve primary outcomes of parenting efficacy, cognitive appraisal of the child and parent–child relationship, improving relevant knowledge, meeting family needs, and child behavior. On the basis of this theoretical framework, the FMF content, principles, and methods were successfully adapted to the mHealth format, with the addition of unique content and features [26].

Figure 1. Phases of a systematic, user-centered design approach to mobile health intervention development and evaluation.
In the second phase, most standard FMF Program content was preserved, but the flow of content delivery was adapted to be more amenable to self-direction by caregivers. There were additional adaptations from a technological perspective. In the third phase, FMF Connect was implemented by leveraging functionalities offered by modern mobile devices and by now ubiquitous access to the internet and Cloud services, as well as considering evolving technological possibilities and the different ways in which users interact with them. In the fourth phase, the initial interface design and functionalities were further refined through stakeholder input with caregivers in focus groups across 5 US cities [26].

This study focuses on the fifth phase of our user-centered design approach for systematic development and evaluation. This phase involves beta-testing of initial app prototypes followed by qualitative analysis of key informant interviews and data drawn from usage analytics to assess the feasibility of the intervention and guide further app refinements.

In this systematic model, the sixth phase involves pilot testing the intervention and trial procedures to establish the best methods for understanding the app in terms of the intervention process and outcomes. This stage is critical for optimizing the intervention and study methods for a larger-scale randomized controlled trial (RCT). The seventh phase involves a rigorous evaluation of app outcomes (and intervention process) through an RCT. Future studies will discuss findings from these phases of programmatic FMF Connect research.

**FMF Connect: A Novel mHealth Intervention**

The FMF Connect app consists of 5 primary components (Figure 2), which have been previously described in depth [26]. Briefly, the Learning Modules make up the core intervention and comprise 12 modules across three levels of educational content and skill development important in parenting children with FASD and behavioral concerns. In addition to brief educational text, the Learning Modules include exercises for active learning, and videos of real families demonstrating ideas and sharing their experiences. The Library contains additional videos augmenting those in the Learning Modules and fact sheets providing psychoeducation on important additional topics, such as mental health diagnoses, medication, trauma, advocacy, and resources. The Family Forum is a peer-moderated discussion forum where users can connect with others, share joys and challenges, and seek support or advice. Finally, the Notebook allows users to save content and exercises they wish to revisit later, and the Dashboard shows the user’s progress.

This study focuses on the fifth phase of our user-centered design approach for systematic development and evaluation. This phase involves beta-testing of initial app prototypes followed by qualitative analysis of key informant interviews and data drawn from usage analytics to assess the feasibility of the intervention and guide further app refinements.

In this systematic model, the sixth phase involves pilot testing the intervention and trial procedures to establish the best methods for understanding the app in terms of the intervention process and outcomes. This stage is critical for optimizing the intervention and study methods for a larger-scale randomized controlled trial (RCT). The seventh phase involves a rigorous evaluation of app outcomes (and intervention process) through an RCT. Future studies will discuss findings from these phases of programmatic FMF Connect research.

**FMF Connect: A Novel mHealth Intervention**

The FMF Connect app consists of 5 primary components (Figure 2), which have been previously described in depth [26]. Briefly, the Learning Modules make up the core intervention and comprise 12 modules across three levels of educational content and skill development important in parenting children with FASD and behavioral concerns. In addition to brief educational text, the Learning Modules include exercises for active learning, and videos of real families demonstrating ideas and sharing their experiences. The Library contains additional videos augmenting those in the Learning Modules and fact sheets providing psychoeducation on important additional topics, such as mental health diagnoses, medication, trauma, advocacy, and resources. The Family Forum is a peer-moderated discussion forum where users can connect with others, share joys and challenges, and seek support or advice. Finally, the Notebook allows users to save content and exercises they wish to revisit later, and the Dashboard shows the user’s progress.

**Figure 2.** The 5 primary components of the Families Moving Forward Connect mobile health intervention for caregivers of children with fetal alcohol spectrum disorders.

**Learning Modules**
- 12 core modules, 3 levels
- Educational text and audio
- Exercises to practice content
- Animation and video

**Family Forum**
- Users share ideas, ask questions, get support
- Organized in subforums
- Moderated by trained peers

**Notebook**
- User builds personalized section for later reference
- Exercises about child, selected content, tools, notes

**Dashboard**
- Summary of progress
- Badges earned, child behavior, ratings, usage metrics

**Library**
- Lists of books, websites, other resources
- Optional fact sheets
experiences and perspectives on the FMF Connect intervention. The feasibility trial was guided by the following objectives, informed by the Eldridge et al [27] conceptual framework:

1. Examine app usage data and crash reports to identify the required technological and functional refinements of the FMF Connect app.
2. Conduct focus group and individual interviews with participants to evaluate the user experience of the FMF Connect app to guide further app refinements.

The study results highlight important directions for the ongoing refinement of the FMF Connect app. By operationalizing a systematic model of app development and evaluation in this project, the findings also have broader implications for mHealth applications. Overall, this project offers generalizable ideas about methods for enhancing the acceptability and rigor of mHealth applications, a vital consideration as the field of digital health rapidly expands.

Methods

Study Design

This study was designed to assess the feasibility of initial prototypes of the FMF Connect intervention from both technological and user experience perspectives to guide further development of the app. This study involved two rounds of beta-testing, which allowed the examination of iterative feedback. The first round of beta-testing (BT1) was conducted from April to May 2019 and included the iOS prototype. The second round (BT2) was conducted from November to December 2019 and an updated iOS prototype and a new Android prototype with the same content and features were tested. Each beta test lasted approximately 6 weeks and included caregivers of children with FASD and providers working with this population.

To address the study objectives, this study used a concurrent quasi-mixed-methods design with equal priority given to both method types [28]. In other words, both quantitative and qualitative analytical methods were used. However, these methods were used to answer different aspects of the research question (ie, the feasibility of FMF Connect intervention prototypes). Therefore, deliberate integration of findings during the interpretation of results was not warranted [29]. After each 6-week trial, focus group and individual interviews were used to elicit participants’ perspectives about the app (qualitative data). Usage data and crash reports were also collected within the app and used to assess the functionality of the app (quantitative data). All study procedures were approved by the university’s institutional review board before initiation.

Participants

A total of 63 participants (45/63, 71% caregivers; 18/63, 29% providers) were enrolled as described below by participant type. Table 1 describes the demographic information. Participants resided in 18 US states.
Table 1. Participant characteristics (N=63).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Caregiver (n=45)</th>
<th>Provider (n=18)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>42 (93.3)</td>
<td>17 (94.4)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>50.41 (11.33)</td>
<td>45.28 (11.60)</td>
</tr>
<tr>
<td>Range</td>
<td>31-73</td>
<td>28-70</td>
</tr>
<tr>
<td><strong>Average age of children with FASD</strong> (n=64)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>9.70 (3.58)</td>
<td>N/A^b</td>
</tr>
<tr>
<td>Range</td>
<td>4-17</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Caregiver ethnicity, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latinx</td>
<td>3 (6.7)</td>
<td>2 (11.1)</td>
</tr>
<tr>
<td><strong>Caregiver race, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American or Black</td>
<td>4 (8.9)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Asian</td>
<td>0 (0.0)</td>
<td>1 (5.6)</td>
</tr>
<tr>
<td>White</td>
<td>40 (88.9)</td>
<td>17 (94.4)</td>
</tr>
<tr>
<td>Native American or Alaska native</td>
<td>2 (4.4)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific islander</td>
<td>0 (0.0)</td>
<td>1 (5.6)</td>
</tr>
<tr>
<td><strong>Education level, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-school diploma or GED</td>
<td>2 (4.4)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Some college, trade school, or Associate’s degree</td>
<td>13 (28.9)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>12 (26.7)</td>
<td>2 (11.1)</td>
</tr>
<tr>
<td>Master’s degree or higher</td>
<td>13 (28.9)</td>
<td>6 (33.3)</td>
</tr>
<tr>
<td>Doctoral or professional degree</td>
<td>5 (11.1)</td>
<td>10 (55.6)</td>
</tr>
<tr>
<td><strong>Relation to child, d n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biological parent</td>
<td>1 (2.2)</td>
<td>N/A</td>
</tr>
<tr>
<td>Adoptive parent</td>
<td>32 (71.1)</td>
<td>N/A</td>
</tr>
<tr>
<td>Foster parent</td>
<td>1 (2.2)</td>
<td>N/A</td>
</tr>
<tr>
<td>Grandparent</td>
<td>9 (20.0)</td>
<td>N/A</td>
</tr>
<tr>
<td>Other relative</td>
<td>2 (4.4)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Family income (US $), n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15,000-24,999</td>
<td>2 (4.4)</td>
<td>N/A</td>
</tr>
<tr>
<td>25,000-34,999</td>
<td>3 (6.7)</td>
<td>N/A</td>
</tr>
<tr>
<td>35,000-49,999</td>
<td>4 (8.9)</td>
<td>N/A</td>
</tr>
<tr>
<td>50,000-74,999</td>
<td>7 (15.6)</td>
<td>N/A</td>
</tr>
<tr>
<td>75,000-99,999</td>
<td>4 (8.9)</td>
<td>N/A</td>
</tr>
<tr>
<td>100,000-149,999</td>
<td>12 (26.7)</td>
<td>N/A</td>
</tr>
<tr>
<td>150,000 or more</td>
<td>9 (20.0)</td>
<td>N/A</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>4 (8.9)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Community type, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>7 (15.6)</td>
<td>N/A</td>
</tr>
<tr>
<td>Suburban</td>
<td>33 (73.3)</td>
<td>N/A</td>
</tr>
<tr>
<td>Urban</td>
<td>5 (11.1)</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Caregivers

Caregivers, including biological, foster, adoptive, and relative caregivers (Table 1), were recruited through multiple mechanisms. Information about the study was shared with providers affiliated with the Collaborative Initiative on FASD and with moderators of national and regional FASD listserves and support groups, to be widely distributed to interested families. We also contacted eligible families in our university’s FASD research registry. Caregivers reported learning about the study from the following sources: provider referral (n=16), FASD research registry (n=3), national and regional FASD listserves (n=9), online support groups (n=5), and nonspecified (n=12). Caregivers were eligible for this study if they had a child with FASD or PAE between the age of 3 and 17 years and lived in the United States. Although FMF Connect is designed for caregivers of children aged 3-12 years, caregivers of adolescents (aged 13-17 years) were also included (n=9). These caregivers were able to reflect on their experiences in parenting their children across the full age range targeted by the app and evaluate the app in this context. A subsample of caregivers (n=7) who had previously completed the standard FMF Program was specifically recruited for this study. These caregivers could offer important insights on what it is like to learn this content in a self-directed manner through FMF Connect versus their prior lived experience of participating in the in-person, therapist-led standard FMF Program.

Providers

Although providers serving children with FASD and their families (eg, medical and mental health providers, occupational therapists, and advocates or educators) are not intended to be direct consumers of the FMF Connect intervention, there were several important reasons to solicit their feedback. First, many serve a diverse range of families and could offer insights to augment those provided by caregivers enrolled in the study. In addition, providers are likely to be a primary future referral source for the FMF Connect app. Gaining their perspective early in development may facilitate app acceptability so that providers will more likely share it with families once it is widely available. In this study, providers working with children with FASD were purposefully sampled through known provider networks relevant to this population. Providers were eligible for this study if they served children with FASD and their families and worked in one of these professions: medical provider (5/18, 28%), mental health providers (8/18, 44%), occupational therapists (2/18, 11%), and FASD advocates or special educators (3/18, 17%). A subsample of providers with experience delivering the standard FMF Program (9/18, 50%) was specifically targeted for this study.

Procedures

Interested participants were sent the study consent form and demographic questionnaire. Informed consent was then completed with the study coordinator over Zoom (Health Insurance Portability and Accountability Act–compliant) or via phone. Participants returned the signed consent and demographic form before receiving the app prototype and installation instructions.

During the 6-week beta tests, participants could use the app at their discretion. As part of the intervention, the participants received weekly emails. These highlighted specific app features and content and provided information on technical assistance access. The Family Forum was moderated by 2 experienced caregivers who had previously completed the standard FMF Program and were supervised weekly by the first author. The study team monitored use and metrics throughout each trial.

---

### Table 1: Characteristics of Caregivers and Providers

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Caregiver (n=45)</th>
<th>Provider (n=18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experience with standard FMF Program, n (%)</td>
<td>7 (15.6)</td>
<td>9 (50.0)</td>
</tr>
<tr>
<td>Round of beta-testing, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta test 1</td>
<td>21 (46.7)</td>
<td>17 (94.4)</td>
</tr>
<tr>
<td>Beta test 2</td>
<td>24 (53.3)</td>
<td>1 (5.6)</td>
</tr>
<tr>
<td>Operating system, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>iOS</td>
<td>36 (80.0)</td>
<td>17 (94.4)</td>
</tr>
<tr>
<td>Android</td>
<td>8 (17.8)</td>
<td>1 (5.6)</td>
</tr>
<tr>
<td>Comfort with technology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>5.73 (1.32)</td>
<td>5.83 (0.92)</td>
</tr>
<tr>
<td>Range</td>
<td>1-7</td>
<td>4-7</td>
</tr>
</tbody>
</table>

*Some caregivers had more than one child with fetal alcohol spectrum disorders.

**N/A: not applicable.

*GED: General Educational Development.

*A total of three biological parents participated in the study. Two participants were biological parents but were noted in other categories (eg, grandparent, who was also a biological parent).

*FMF: Families Moving Forward.

*For caregivers, experience with FMF denotes completing the FMF Program, and for providers denotes training and delivering the FMF Program.

*Comfort with technology was measured using a Likert scale ranging from 1 (low) to 7 (high).
Bugs and crashes were tracked, and updates were sent to address problems or expand functionalities. Following each 6-week trial, participants were asked to complete individual or focus group interviews with a member of the study team. In BT1, focus groups were organized by participant type and usage pattern (eg, number of modules completed, relative time spent in the app). Interviews were offered to participants when focus group participation was not possible because of schedule conflicts or comfort levels. For logistical reasons, planned individual interviews were conducted with all BT2 participants. Data collection was completed via Zoom for all but one BT2 caregiver participant, who preferred an in-person interview to better accommodate hearing loss. The questioning route (details provided in Multimedia Appendix 1, Table S1) was similar across both beta tests. However, two topics were added to BT2 to assess participants’ perspectives on new features. Topics included Global Impressions & Experiences; Usage/Engagement; Technology; Utility; and Experience with Individual Components (eg, Learning Modules and Family Forum). After introducing the Global Impressions & Experiences topic, interviewers were given flexibility regarding the order in which they covered subsequent topics. This was done to facilitate conversational flow and follow the participant’s lead during the discussion.

Data Analysis
Audio and video recordings of individual and focus group interviews were transcribed by the research staff. Observations of nonverbal (eg, tone, affect, referencing app on phone) were integrated within each transcript. All transcripts were checked for accuracy and completeness. The data were then imported into Atlas.ti for coding and analysis. Four research team members conducted primary analyses: one of the principal investigators, a graduate student, and 2 research staff. All members of the analysis team were involved in data collection.

A thematic analysis [30,31] was undertaken to understand participants’ experiences using the app from both technological and content standpoints. Coding methods were selected a priori to inform further app refinements for subsequent larger-scale trials. These include structural, evaluation, and value coding [31,32]. Structural coding was used to delineate when participants discussed different app components (eg, Learning Modules and Family Forum). Evaluation coding was selected to identify participants’ positive or negative judgments about the FMF Connect app and recommendations for further improvements. Value coding was used to identify caregiver values, attitudes, and beliefs related to the experiences of raising a child with FASD and using the FMF Connect app. For provider data, value coding was only used when providers spoke about their perceptions of the values, attitudes, and beliefs of caregivers.

Systematic thematic coding of transcripts was completed between May and December 2020. Four parent interviews from BT1 were randomly selected and independently coded by line by line all 4 coders to establish the study codebook. Weekly meetings were held to establish consensus and operationalize first-level codes. The remaining BT1 caregiver interviews were then distributed across coders, taking care not to assign transcripts to the team members who had conducted the interviews. Weekly coding meetings were held to address any coder questions or suggestions for new codes.

Following completion of BT1 parent interviews, the team engaged in code mapping to organize and consolidate first-level codes into preliminary second-level pattern codes to facilitate subsequent coding [31,32]. BT1 provider coding and BT2 coding followed the same process. The preliminary second-level pattern codes represented the data well across BT1 providers and all participants in BT2. No new second-level pattern codes were added across these participants, suggesting adequate data saturation and consistency across trials.

Participant matrices [31] were used to examine variance in second-level pattern codes across participants and several key demographic features (eg, prior participation in FMF, BT1 vs BT2). Participant demographic variables were imported into Atlas.ti, and code co-occurrence tables were examined to assist with this process. Team members iteratively consolidated and interpreted the connections among the data through analytic memo writing to derive the final analytic model.

App usage metrics were examined for caregivers. Usage data were extracted from the cloud services used for the app. Descriptive statistics were calculated for several indices (eg, number of modules completed, number of posts in the Family Forum, and total time spent in the app). Learning Module completion patterns were examined using graphical methods.

Results
Objective 1: Examine App Usage Data Metrics to Identify Any Needed Functional Refinements to FMF Connect

Overview
Table 1 provides the breakdown of participants and the type of operating system across the beta tests. A total of 84% (38/45) of parents (BT1=16/20; BT2=22/25) and 94% (17/18) of providers (94%; BT1=16/17; BT2 1/1) installed the FMF Connect app. In BT1, 4 updates of the iOS app were distributed; in addition to bug fixes and performance improvements, updates included the ability to see if there were new posts in the Family Forum since the last user’s login, the addition of the Profile Graph Tool, and improvements in the screen unlocking experience and avatar customization. In BT2, 3 updates of the iOS app and 2 updates of the Android app were distributed for bug fixes and performance improvements.

On use days, caregivers averaged approximately 20 minutes (mean 19.63, SD 19.59 minutes) using the app. The largest amount of time was spent watching videos in the Learning Modules (45% on average). In the Family Forum, there were 54 original posts in BT1 and 45 posts in BT2. A total of 67% (30/45) of users posted at least once in the forum. Not unexpectedly, patterns of use varied considerably among caregivers. The standard FMF Program is similar in total time spent on other parent training programs. The FMF Program typically involves 6 to 9 months of therapist-delivered content in a collaborative therapeutic relationship with caregivers.
(sessions every other week). Therefore, we did not necessarily expect users to complete all 12 modules in the initial 6-week test. Usage data differed according to the operating system and will be discussed separately.

**iOS Usage**

Figure 3 shows the number of Learning Modules completed by iOS users by beta test. A total of 31% (10/32) of iOS users who installed the app completed at least through module 6 (an average of 1 module per week). We also examined the time spent on each activity within the modules. Bar graphs for each module were created with minutes spent in sections by the user (not shown). Through visual inspection, 5 usage tiers were characterized based on time devoted to activity completion and conceptual organization of modules, ranging from tier 1=higher robust use to tier 5=installed but no module use (Table 2 provides descriptions and number of users per tier). Figure 4 shows this classification graphically. Graphs were also created for Learning Module total time per day to illustrate usage patterns by tier (Figure 5).

**Figure 3.** The number of Learning Modules completed by caregivers using Families Moving Forward Connect on iOS phones by beta test. The first round of beta-testing (BT1) had 4 caregivers who did not install the app and 1 who installed but had no module completion. The second round of beta-testing (BT2) had 2 caregivers who did not install the app and 2 who installed but had no module completion.
**Table 2.** Learning module usage tiers for iOS users who installed the Families Moving Forward Connect app.\(^a\)

<table>
<thead>
<tr>
<th>Tier</th>
<th>Description</th>
<th>iOS users (n=32), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Higher robust use:</td>
<td>6 (19)</td>
</tr>
<tr>
<td></td>
<td>• Completed at least up through module 9 (level 3) or finished all modules</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Generally adequate time to complete activities</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Moderate use:</td>
<td>4 (13)</td>
</tr>
<tr>
<td></td>
<td>• Completed at least up through module 6 (level 2) with adequate time to complete core activities</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• May have had some variable usage (eg, skipping through activities) in some sections or modules after 6</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Good level 1 use, but drop off:</td>
<td>8 (25)</td>
</tr>
<tr>
<td></td>
<td>• Demonstrated adequate time to complete activities in modules 1-3 or 1-4 (level 1)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Minimal or low use:</td>
<td>11 (34)</td>
</tr>
<tr>
<td></td>
<td>• Only completed up through modules 1 or 2, or</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Inadequate time to review information or complete activities (skipped through screens)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Installed but no module use</td>
<td>3 (9)</td>
</tr>
</tbody>
</table>

\(^a\)6 iOS users did not install the app; 4 were in first round of beta-testing, and 2 were in the second round of beta-testing.

**Figure 4.** Tier classifications of usage for iOS users of the Families Moving Forward Connect mobile health intervention.
Figure 5. Total time spent in Learning Modules by day of use for iOS users grouped by usage tier.

Tier 1 users tended to distribute their use of the app across regular sessions of 15-30 minutes or several very lengthy sessions of 60-120 minutes toward the beginning of the trial; total app use time was between 4 and 8 hours for these users. Tiers 2 and 3 users tended to use the app more sporadically and for a shorter period than those in tier 1. A few users also had a spike of high use toward the end of the trial, which likely coincided with scheduling a study interview. Many tier 4 users logged in just once or twice. The 2 tier 4 users with a higher percentage module completion quickly skipped through most screens.

iOS usage tier membership was also examined by participant characteristics through visual inspection and statistics (chi-square and analysis of variance). Usage tier did not differ by caregiver age ($F_3=1.41; P=26$), level of education ($\chi^2=6.15; P=.73$), previous receipt of standard FMF ($\chi^2=2.77, P=.43$), target child age ($F_3=2.46; P=29$), or income level ($\chi^2=23.17, P=.18$). Tiers varied according to comfort with technology ($F_3=9.07, P<.001$). Participants who rated themselves lower on comfort with technology were surprisingly more likely to be in tier 1.

Android Usage

Given the small size of the development team, the Android prototype was implemented on a schedule deliberately set behind the iOS prototype, and not all functionalities were available at the time of beta-testing. Nevertheless, the app was distributed to Android users for the initial testing of the technology. Usage patterns for Android users in BT2 differed from those of iOS users across tests, likely because of technical problems with the Android prototype. Incomplete data appear to have been recorded in cloud services for Android users. This may have occurred because of synchronization issues after users completed modules offline. The data recorded show that 3 of the 6 users completed at least some (but likely all) of each of the first 4 modules. The other 3 users only have data recorded for some of module 3. Toward the end of BT2, a few users reported that they could not unlock modules in level 2 of the app. It is possible that these users would have proceeded further in using the app if they had not encountered this technical barrier.
Objective 2: Using Qualitative Methods, Evaluate the User Experience of FMF Connect to Guide Further App Refinements

Overview

We attempted to interview all study participants regardless of whether they installed the app or their usage pattern. Focus group and interview data were available for 26 parents (BT1: 3 in focus groups, 7 interviews; BT2: 16 interviews) and 16 providers (BT1: 6 in focus groups; 9 interviews; BT2: 1 interview). Of the 19 caregivers not interviewed, nearly all had little to no app use (1 in tier 2; 8 in tier 4; 3 installed but no use; 7 did not install). Themes did not generally differ between iOS and Android users, with the exception of ease of use and technological problems noted below.

General comments on the app were the most frequent (166 coded segments). Following this, the Learning Modules (151 coded segments) and Videos (106 coded segments) garnered the most discussion across participants, with often lengthy, detail-rich segments. The Family Forum (94 coded segments) and Library (59 coded segments) received a modest amount of discussion, with the Notebook (32 coded segments), Dashboard (32 coded segments), and Logo/Icon (11 coded segments) eliciting limited discussion of short duration.

Findings From Values Coding

Although not explicitly queried in the questioning route, caregivers often spoke of their attitudes, beliefs, and values. These provide an important context for understanding their evaluation of the FMF Connect intervention. Themes did not vary by round of beta-testing, Learning Module usage tier, smartphone operating system, caregiver type (adoptive vs relative vs biological), child age (child vs teen), or previous standard FMF Program experience.

FASD are often described as complex and confusing. For example, one caregiver (FG028) shared the following:

I remember the early days and thinking, “What on God’s green Earth, you know, is wrong with this child? What is going on?” ... “You could do this yesterday, what do you mean you don’t know where your shoes are? (Voice raises) How do you put your shoes on the wrong feet 90 percent of the time? How does that happen (laughs)?” ... the behavior is just baffling in the beginning.

The complexity of FASD is further complicated by the fact that many children with FASD have experienced trauma and have other comorbid conditions, as illustrated by a caregiver of teenagers (FG079):

That was our struggle with our kids when they were little. Is it because of their alcohol exposure? Is it because of the trauma? Is it because of who they are? Is it a mental health thing? And, you know, everyone has their own opinion when you take them to therapists and doctors.

Participants emphasized that FASD information and resources are often lacking, which is associated with feelings of frustration, grief, and being overwhelmed. For example, one provider (FG059) stated as follows:

So many parents are desperate for answers, they’re desperate for information. ... There’s a lack of resources and lack of evidence-based intervention in most communities.

A caregiver (FG043) also emphasized difficult emotions arising from inadequate supports:

There is great remorse and guilt... I had a child with FASD because no physician took the time to get me into treatment when it was very obvious that I needed treatment.

Parents further described FASD as isolating. For example, a caregiver (FG049) described the following:

I don’t have the opportunity to talk to other parents... umm... ever (laughs) who have children with FASDs, umm, so that is very isolating. ... Because we raise their volume can’t even find a doctor who knows what they’re talking about, let (normal volume) alone, umm, another grown-up going through it.

Caregivers spoke to their desire to do anything to help their children be as successful and independent as possible. Given the limited number of knowledgeable and skilled professionals, this often results in the need to educate others about FASD. For example, a caregiver (FG065) explained as follows:

We, as parents, you’re always educating other people. And so teachers and parents, some doctors... anybody working with your kid, you know, occupational therapist, speech therapist, therapist... (emphasis) A lot of people do not understand.

As a result, caregivers raising children with FASD highly value access to information about FASD, people who understand their experiences, and the ability to connect and share resources with others. Participants expressed the belief that mHealth interventions, such as FMF Connect, are needed to help address barriers. For example, a caregiver (FG071) stated as follows:

We don’t have anything to really go to, so I think it’s really great to have the educational piece but also to have the forum, to kind of link people together because you do feel really isolated.

Many related these barriers to strong emotions, such as frustration. One caregiver (FG084) described the following:

A lot of people go on here to learn things. But, to be honest, (strong emphasis) I think most people go on the apps and go on the groups just to be with other people who are going through it... nobody in my life understands (frustrated tone)... And if I post in there, it would be mostly just to exahle to other people who get it, you know? And for somebody else to come on there and say, “I get it.” You know? (slow, normal volume) It’s just sometimes, that’s all you need for the day. Is for somebody to say, “I get it.”

Because caregivers raising children with FASD are often overwhelmed, participants articulated the need to use their

https://formative.jmir.org/2021/12/e29687

 JMIR Form Res 2021 | vol. 5 | iss. 12 | e29687 | p.113

(page number not for citation purposes)
limited time wisely and valued choice and autonomy in self-directed learning. A provider (FG040) stated as follows:

You know, people are busy and they just-they want to know how much time these are going to take...knowing how much time you generally might need to spend with something... just might keep people engaged.

A caregiver (FG051) emphasized the value of choice:

I think the more people can make choices in what they’re doing, the more the buy in and (laughs) you know, the more likely they’re gonna do it and, and be happy about doing it.

Evaluative Coding: Positives

Across participants, there were 2.5 times as many positively evaluated coded segments as negatively evaluated segments. The vast majority of themes were consistent across demographic and usage tier variables. The few differences are discussed in the relevant sections.

Global Impressions

Global impressions of the app (eg, “I love it!” “It’s wonderful,” and “I enjoyed it.”) were nearly all positive (93 global positive vs 1 global negative segment). Participants appreciated the accessibility of the app and how they could fit it within their everyday lives. For example, a caregiver (FG015) said:

This makes it easy for me, it’s right there at my fingertips.

A provider (FG032) also highlighted the benefits of FMF Connect as a smartphone app:

The majority of my families do not have a home computer this year. The majority of them do everything off of their cell phone so that’s their only access to the internet.

Most of the iOS users across both beta tests also commented that they found the app easy to use. In contrast, most Android users did not mention this theme, likely because of technological problems in the Android prototype. The participants also made positive comments about the app’s appearance.

Learning Modules and Libraries

Positive evaluations of the informational content provided by the app had the most coded segments across all codes (139 segments). In fact, every parent made at least one positive comment regarding the informational content of the app. Participants appreciated how the FMF Connect app made this information more accessible to them. For example, a caregiver (FG056) explained:

It’s hard to find good information on FASD, and I thought that it was kind of cool that it was on my phone, all together, in one spot.

Several participants spoke about the quality of information. For example, a caregiver (FG065) stated:

I thought it was all very relevant and research based which I appreciate (laughs) very much.

Informational content also overlapped with parents’ values of understanding their children. Several parents provided specific examples of how app content helped them better understand their children’s behavior and respond differently. Participants were especially enthusiastic about the ability to share information from the app, particularly with teachers (relating to the themes of needing to educate others and value of sharing resources). For example, a caregiver (FG079) described as follows:

I printed out something to take to her meeting that I have next week for the teachers. ... I think that was that was the big thing that I was excited to find this stuff to give to them.

Participants also spoke positively about aspects of the videos, including diverse representation of families, range of child ages and degree of problems, and specific ideas and strategies to try. Caregivers especially appreciated that the videos featured real families, as illustrated by a caregiver (FG056):

I was like, really excited when I first started and I was watching the videos and I was like “Oh my gosh!” You don’t get to see how other FASD kids live and how they are, so it was really cool to see like real families and real kids. Like, that was my personal favorite part of the whole thing.

Positive comments about the videos also often co-occurred with themes of parents feeling less isolated and validated in their experiences as parents. For example, a caregiver (FG052) shared the following:

[The videos] kept me grounded and mindful, umm, that, first I’m not in this alone. And other people are experiencing the same thing, and here are some things that they found that worked.

The exercises within the Learning Modules received positive evaluations by some users. Some participants commented on how the various exercises and games also helped them reflect on learning content and apply information to their children.

Similar to feedback revealed in prior studies of FMF Connect [26], the step-by-step progression of access to content in the Learning Modules and Library received mixed evaluation. Discussion of this theme was often intertwined with participants’ previous knowledge and experience with FASD and thoughts about whom the app is best suited. Every provider, especially those trained in the standard FMF Program, spoke to the need for and positive aspects of the step-by-step progression of these components. Most caregivers also spoke of the advantages of step-by-step progression. A caregiver (FG050) stated:

I liked how the progression went. I thought it was easier to be able to focus and break it down and think about that particular section at a time.

Several parents emphasized how this progression made learning less overwhelming. Although less enthusiastic about the step-by-step progression for themselves, more experienced parents felt this would be very beneficial for parents of newly diagnosed children and thought the app was most well suited for this group. For example, a caregiver (FG028) stated as follows:

https://formative.jmir.org/2021/12/e29687

JMIR Form Res 2021 | vol. 5 | iss. 12 | e29687 | p.114

(page number not for citation purposes)
I’ve been in the trenches for many years (laughs). It’s nothing new to me. And I know-like, I guess I kept thinking as I was using it, this would be super helpful for someone who was in my shoes 8 years ago.

Participants also spoke to the benefits of the FMF Connect app as a refresher. They described how parents could go back at any time and review content and apply it to new challenges. For example, one caregiver (FG061) stated as follows:

I have thoroughly enjoyed all of the little activities we had to do that reinforced everything... Being able to go back to see it over and over again. To get it going in your little brain “Ok, I can, I can do this.”

Family Forum

Participants were positive about the inclusion of the Family Forum and articulated its potential to reduce isolation and help parents connect with others who have shared experience. Evaluations of the Family Forum often overlapped with parents’ values of connection and people who understand. For example, a caregiver (FG025) explained as follows:

I thought it was useful that people could ask questions, like the real issues that we deal with. And get some kind of advice and some kind of help because I find that we deal with things that most people aren’t dealing with all the time.

Participants particularly liked that the Family Forum was moderated and that there was a special section where their posts were saved for later reference.

Dashboard and Notebook

The Dashboard and Notebook received fairly limited discussions and were primarily associated with nonspecific positive impressions. However, during beta-testing, these components had relatively limited functionality. In BT2, a Tip of the Day feature was added through a push notification that subsequently appeared on the Dashboard for that day. Enthusiasm was high for this feature, and nearly every participant in BT2 provided a positive evaluation. For example, a caregiver (FG042) explained as follows:

The constant tips - I mean it’s like having a social worker right in your home with you all the time ... It doesn’t matter if you’re having a good day or a tough day, having that positive reframing and, it, it’s like a breath of fresh air, it’s like, ok, slow down, you know this, and here’s a reminder, yes, ok, (laughs) you have to let go of that and you have to do what the tip says.

Participants also described how the Tip of the Day was useful in reminding them to use the app.

Motivators and Facilitators of App Use

Participants identified app content and the ability to connect with others as primary motivators of wanting to use the app. One caregiver (FG072) described this as follows:

That’s a big motivator, too, is just wanting to have one more tool, I have the books, I’m watching the YouTube videos. I’m doing everything that I can do. But I have my phone with me all the time.

Another caregiver (FG068) shared as follows:

I think for a lot of parents the motivation [to use the app] would be just, you know getting help...And connect with others...Cause it’s hard when other people don’t understand.

Parents described using the app most often at night once their children were in bed or during moments of downtime. For example, one caregiver (FG050) used the app:

Whenever I had free time. Usually, before the kids woke up, or after they went to bed. So it was just, whenever I had time, I would do it for a couple minutes here, a couple minutes there.

The parents who progressed the furthest in the Learning Modules (ie, tier 1) described strategically planning ahead for manageable segments of time to work through the app. One caregiver (FG061) described this as follows:

I’d spend at least thirty minutes every day, if not an hour, if I had the time, I’d make sure I had the time, but not everybody has my schedule.

Evaluative Coding: Negatives

Overview

Negative evaluations comprised 28.36% (312/1100) of the total number of evaluation segments coded, so the study methods were successful in eliciting these. These negative evaluations largely emphasized technological issues, constructive feedback relative to navigating the app, and barriers to use. The only negative global impression segment across all participants was from an Android user (FG066) who experienced difficulties loading the videos and was disappointed by the level of activity in the forum:

Um, (laughter) to be honest I wasn’t very fond of [the app]... Um, granted there were very few people on as testers, but... I didn’t see a whole lot of conversations going on. Um, its, the videos themselves, half the time they didn’t work for me.

Technological Problems

Some users experienced technological problems using the app. These were the most significant for iOS tier 4 and Android users. BT1 included all iOS users. In BT1, some participants experienced confusion or difficulty with TestFlight (an iOS app that allows beta-testing before release in the App Store). For these users, difficulty with TestFlight impacted the initial installation or updating the app. Several updates were released early in BT1 because of some crashing and videos not loading consistently. In addition, several participants reported some difficulties in saving their progress in the Learning Modules; after refinements, this was not an issue reported in BT2.

Android users (who were only part of BT2) described more significant technological problems that resulted in barriers to using the app. These issues included app freezing, some inaccessible links, and problems loading videos or unlocking later content. For example, one caregiver (FG066) described this as follows:

https://formative.jmir.org/2021/12/e29687
In this app, it was just stuck. I couldn’t get out of it, and I couldn’t do anything unless I completely closed it down.

A provider (FG036) reported problems with video loading:

I could not get the first video to play. Umm, and that was something that had happened throughout, like me trying to watch the videos is they just keep buffering and buffering.

Navigation: Family Forum

On the basis of previous feedback regarding the design of the FMF Connect app [26], the Family Forum was initially laid out

On the basis of this feedback, the Family Forum interface was redesigned for BT2. Subsequently, there were fewer negative evaluations related to navigating the Family Forum (Figure 6B). In BT2, negative evaluations mainly centered on parents not noticing or understanding the tags or categories at the top of the screen.

Participants also noted that the overall use of the Family Forum was lower than expected, and conversations tended to stagnate because of insufficient notifications. For example, a caregiver in BT1 (FG051) commented as follows:

I went in and I introduced myself, and I was excited about the forums and connecting with other people but then, I think having so many different forums and not really getting notifications of when people were posting, umm, that made me just keep forgetting about them.

Navigation: Learning Modules and Library

As mentioned previously, the step-by-step progression of the Learning Modules received a mixed evaluation. Reactions to move through the content in order varied by usage tier and previous experience with FASD. Some experienced users, with

robust tier 1 use, identified that content was redundant for them. However, none of these experienced users described their experience as tedious or found that step-by-step progression was a barrier. These themes were only present for lower-use tiers 2 to 4. Negative evaluations related to step-by-step progression occurred more often for participants with a greater degree of previous knowledge and experience with FASD. One caregiver (FG066) expressed this sentiment as follows:

I’m just simply not going to sit through you know, 5, 10 hours whatever it is of information and watch it and everything when I already know it just to get to something I don’t know. There are better ways for me to do it.

Similarly, about the Library, a caregiver (FG069) explained as follows:

Maybe it’s because I already came in with a fair amount of knowledge, but I wasn’t a big fan of different things opening up as I went. I would have preferred to have jumped in and seen everything immediately.
Participants in BT1 found the number of videos and screens per module to be overwhelming and a barrier to completion. For example, one caregiver (FG051) explained as follows:

I’d learn about something and then there would be people sharing their experiences, which is great. But after the first two or three I was like ok, I get it, and then there were like 12! (laughs)... It took more time than I maybe had at the moment, and I wanted to kind of get past the videos and work on whatever was next.

Given this feedback, refinements were made for BT2, which included a table of contents for each module with fewer screens per section. A smaller number of videos were customized for users on the basis of user-imputed data (eg, child age and behavior problems), and the remaining videos were stored in the Library for further viewing. These changes were very positively evaluated by the BT2 participants.

Barriers

Identified barriers to app use generally corresponded to the negative evaluations discussed above. Time and attentional resources were also described as barriers. One caregiver (FG062) explained as follows:

Basically I have free time for like 10-15 minutes at night when I’m putting the youngest kids to bed. And of that, I have very few minutes where I can actually listen.

Another caregiver (FG071) also commented as follows:

I think for me it was too much work at that time of the day... A big issue for me is knowing my energy level at that time of the day and knowing what I had to do.

Providers offered additional insights into potential barriers on the basis of their experiences working with families raising children with FASD. Several providers mentioned factors, such as age, literacy level, English as a second language, and comfort with technology as potential barriers for some families they work with. For example, a provider (FG047) stated:

I have other [patients] that are great-grandmas and grandmas who barely have a computer and have a cell phone mostly to accept phone calls, and make phone calls and “I don’t know about these apps honey I don’t want to deal with that.” And then of course a big barrier here ...is we have a big Spanish speaking component.

Another potential barrier raised by providers was parents feeling intimidated or lacking confidence in implementing strategies demonstrated by parents in the videos. For example, a provider (FG032) described as follows:

As I’m watching the videos, I know some of the parents I’m working with would be petrified to just watch [child’s name]’s mom because they’re like, “I could never do that. I have six children. How is this going to work?”

Evaluative Coding: Recommendations

The participants offered a number of useful recommendations to improve navigation and enhance engagement with the app (Table 3). Some of these recommendations were directly related to aspects that were negatively evaluated. As mentioned above, these led to refinements to the app in between trials, and additional changes are underway.
Table 3. Recommendations offered by participants for further refinement of the Families Moving Forward Connect mobile health intervention.

<table>
<thead>
<tr>
<th>Category</th>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>General app-wide functions</td>
<td></td>
</tr>
</tbody>
</table>
| Engagement features                   | • More robust notification system  
|                                       | • Coaching                        
|                                       | • Changes to weekly emails        
|                                       | • Tip of the day                  |
| Navigation—making things easier to find or use | • Search function                  
|                                       | • Overlays and tutorials           
|                                       | • Navigation shortcuts            
|                                       | • Multiple platforms (ie, phone, computer, tablet)                            |
| Broader access                        | • Spanish language, closed captioning                                           
|                                       | • Functions to allow consideration of multiple children with FASD within the app| 
|                                       | • Ability to share the app and other materials with others                    |
|                                       | • Companion apps for teachers, children/teens, and others                      |
| Learning Modules and Library          |                                                                               |
| Navigation/organization               | • Table of contents for each module, with fewer screens per section            |
|                                       | • Open access to all content from the start                                   |
|                                       | • Consolidate or offer selected number of videos                             |
|                                       | • Fewer clicks to start videos                                                |
|                                       | • Live links to other web sites                                               |
|                                       | • Having content available on (1) app screens so easier to read and (2) PDF for easy sharing |
| Content                               | • More real-life videos and practical strategies                              |
|                                       | • Research summaries                                                           |
|                                       | • Information specific to birth parents and ways to deal with guilt, grief, shame, and stigma |
|                                       | • Additional ideas for self-care                                              |
|                                       | • Tips for advocacy and navigating systems                                     |
|                                       | • Strategies for facilitating social interactions/friendships                  |
|                                       | • Video examples of clinicians working with parents                           |
| Family Forum                          | • Different forum interface                                                    |
|                                       | • Open access to subforums                                                     |
|                                       | • Discussion starters                                                          |
|                                       | • Regional or state subforums; Provider directories                           |
|                                       | • Ability to direct message other users                                       |
| New features under development        | • Behavior tracker                                                             |
|                                       | • Coaching                                                                     |
|                                       | • Daily ratings                                                                |

aFASD: fetal alcohol spectrum disorders.

Additional recommendations were closely related to the values expressed by the participants. For example, caregivers emphasized the importance of using their limited time wisely and having choice and autonomy over their self-directed learning. As a result, they recommended open access to all content in the app from the start and tools to make it easier to refresh their learning when they needed it. For example, a caregiver (FG057) stated as follows:

I understand the whole idea with the yellow brick road, I think that’s great, but I think for me, I would like things that I could just tap on that road, to kind go back and forth in some other groups and videos and stuff, and kinda jump around a little bit more...the app was just leading down a one road, which is great, but sometimes I like to take the detour.

Another caregiver (FG055) spoke to the benefit of repetition and refreshing her learning:

I think it would take a lot of repetition... for me to benefit the most from all the exercises. So, I think it would be nice … to do the same exercises over and over and over and over again. Especially when the behavior has just happened, and I want to go back and I want to do that exercise for that behavior.

Caregivers also value resources to help them better understand their children. Most caregivers liked the idea of a behavior
tracker to help them monitor and notice patterns in complex behavior problems.

A number of recommendations related to a cluster of attitudes, beliefs, and values; specifically, caregivers’ need to understand their child, educate others about FASD, and connect with people who understand. For example, the recommendation for state or regional subforums in the Family Forum would help caregivers rely on others in their area to identify and share available resources and connect locally. One caregiver (FG049) stated as follows:

“I have been in the hunt of my life trying to find my daughter just somebody to take her to that has even a basic understanding of her diagnosis. So, if I had like, just a place where I could go and know like, “These are the people from my state.” Like, “Where do you take your child?”

Some participants also liked the idea of having a coach or expert associated with the app who could help them understand FASD, connect them with needed resources, and provide feedback and discussion relevant to their child. They recommended additional content and tools that would make it easier to share information from the app with teachers, providers, and other people working with their children. Some participants went further and recommended separate apps or components specifically for use with their children or with teachers. For example, one caregiver (FG056) said as follows:

“I really liked the material. My daughter’s teacher would benefit from- it would be cool to like, have her access it also.

Another caregiver (FG069) also noted as follows:

“You know what I’d really like to see is, is this program targeted at medical providers!

On the basis of their experience with other apps, caregivers highlighted some additional facilitators that might help them engage more with the app. The most frequent suggestion was a more robust notification system. One caregiver (FG015) described this as follows:

“With so much...going on, with the stress level and all this stuff happening, sometimes you get easily, so overwhelmed you don’t get a chance. If it’s not at my fingertips or not right there, then it’s out of sight, out of mind, you know.

Similarly, a caregiver (FG042) stated as follows:

“I find it helpful when I get a little prompt in my text or in my email, just to remind me that, you know, the app is here and you can, you can just click from that email or that text and jump into the app.

Participants thought it would be best if the user could customize the type and frequency of notifications to their preferences.

Several biological parents emphasized the importance of the representation and education of their experience in reducing stigma. For example, one caregiver (FG084) explained as follows:

“You don’t see too many statistics that talk about successful parenting... by biological [parents], you know what I mean? ...And so, I tried to...I tried to be the best mom I can and I try to show people.

Another caregiver (FG043) described the following:

“Birth parents understand why people are angry and they just want to prevent the next birth mother from drinking. But I think [education about why people continue to drink] would make them feel more welcome because we all understand that the general public doesn’t get alcohol use disorders. ... Having that education, you know, education is the key.

One caregiver (FG084) emphasized the importance of the moderator in creating a welcoming space:

“When I did my introduction, I was a little worried. Umm, the moderators were the only ones who welcomed me and that’s okay. I knew that was probably going to happen... I did post a couple of times... But nobody made me feel unwelcomed... and that is more important.

A caregiver (FG066) also made several specific recommendations for integrating additional support for biological parents:

“Granted everybody’s in [the Forum] together. Maybe there would be an area that (pause) birth moms could go do, specific. That, not necessarily saying they’re any different from the other moms, because everybody’s a mom whether by choice or by birth, but more so, because birth moms often have the shame and the guilt associated that need to be worked through ...If you touched on it in the Learning Modules that would be great because a lot of women we find have a lot of guilt.

Providers also offered recommendations to increase accessibility for a broader range of diverse families, such as closed captioning and speech-to-text. Several providers also recommended Spanish-language features, given the large mono- and bilingual Spanish-speaking population they serve. Finally, both parents and providers mentioned that families often have multiple children with FASD and wanted features within the app to better accommodate this.

Discussion

Principal Findings

This study presents critical stakeholder feedback and usage data from two rounds of beta-testing of prototypes of the FMF Connect mHealth intervention for caregivers of children with FASD. This fifth phase in the systematic user-centered design approach to the development and evaluation process of the app (Figure 1) yielded important insights on the acceptability and usage patterns of FMF Connect. The findings have implications not only for subsequent app refinements specific to FMF Connect but also for broader mHealth and digital parenting and developmental disability-related interventions.
Two primary research objectives were examined to evaluate the feasibility of the FMF Connect intervention during this phase of our systematic approach.

**Objective 1: Examine App Use Data Metrics to Identify Any Needed Functional Refinements to FMF Connect**

First, we considered how well the app worked for diverse users from a technological standpoint. iOS users in both rounds of testing generally described the app as easy to use. Usage patterns were variable, but surprisingly, had few associations with demographic factors or how participants evaluated the app. Users in the “higher more robust use” tier 1 were more likely to strategically set aside time to engage with the app. Technical issues were more significant for users in lower usage iOS tiers. Recognizing technical issues during beta-testing, we released multiple updates in each trial to fix bugs and improve performance.

It should be noted that BT2 also included a new Android prototype that was designed in alignment with the iOS prototype. Despite alpha testing within the development group on several devices and simulators, Android users experienced greater technological difficulties. These included issues with loading the videos, unexpected crashes and issues when accessing some of the Learning Modules, and synchronization issues between the Android app and Cloud storage. Relative to iOS, it is possible that the more limited regulation and decreased consistency among Android devices and supported versions of the operating system contributed to these technological barriers. Consistent with the objective of beta-testing, we expected to identify technological issues in the context of real-world user implementation. This is a valuable part of the process of user-centered design and informs needed functional refinements.

Usage data also highlight the need to carefully consider design features for engagement and operationalize these features for the FMF Connect app. The tier classification of usage patterns in this study showed evidence of nonusage attrition, with 44% (14/32) of iOS users who installed the app with minimal or no use (tiers 4 or 5). Although this is within the range of premature dropout rates observed for in-person parenting interventions in community settings [33], much work is required to improve accessibility and engagement. Research has called for a science of attrition [34], arguing that understanding dropout and nonuse in mHealth interventions is essential to optimize interventions for targeted populations. In line with this, we examined participant characteristics across tiers and found that those with higher comfort with technology were more likely to have low or no use. It is possible that these participants had other supports in place or had already discovered the information on the web. This highlights the likelihood that the FMF Connect app is of particular importance for underserved populations and could help address social disparities.

More general data on usage patterns in mHealth and digital interventions highlight the critical need to carefully consider design features for engagement. On average, approximately 25% of apps are only used once after download [35], and only 29% of app users were still using an app 90 days after download [36]. Overall, these statistics suggest that engagement in self-directed mHealth and digital interventions can be challenging. This is especially true for parents, given the many demands on their time. Participants in this study noted barriers, such as lack of time and feeling exhausted and overwhelmed, which will arise for any parenting intervention. However, this may particularly be true for interventions targeting parents who are faced with the challenge of raising children with disabilities. A portion of the users in this study were able to strategically prioritize time to engage with the app on a regular basis. Clearly, additional features and supports are required to facilitate and maintain engagement for other users.

In a cogent review, Wei et al [37] identified seven themes that improve user engagement with mHealth applications. As shown in Table 4, the participants in this study independently identified and positively evaluated aspects of the FMF Connect app that correspond to each of the 7 engagement themes. Especially strong were features supporting the themes of interface esthetics, message presentation, and credibility. The findings reveal that many existing design features thought to enhance engagement were already built into the FMF Connect app. However, useful recommendations for further refinement relating to the four themes of navigation, personalization, reinforcement, and communication were suggested by participants.
Table 4. Comparison of design features of the Families Moving Forward (FMF) Connect mobile health intervention with thematic checklist of features to improve user engagement.

<table>
<thead>
<tr>
<th>Design feature themes that improve engagement</th>
<th>Existing features in FMF Connect beta tests</th>
<th>Recommended features for future development</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interface esthetics</td>
<td>Pleasing color scheme</td>
<td><em>a</em></td>
</tr>
<tr>
<td></td>
<td>Positive evaluation of graphics</td>
<td></td>
</tr>
<tr>
<td>Navigation</td>
<td>Easy to use (iOS)</td>
<td>Tutorials/overlays</td>
</tr>
<tr>
<td></td>
<td>Table of contents (BT2)</td>
<td>Direct search feature</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Navigation shortcuts</td>
</tr>
<tr>
<td>Personalization</td>
<td>Selected videos (BT2)</td>
<td>Open access to content</td>
</tr>
<tr>
<td></td>
<td>Profile graph</td>
<td>Multiple children</td>
</tr>
<tr>
<td></td>
<td>Exercises—provision of goal setting and feedback</td>
<td>Personalize notifications</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Behavior tracker tool</td>
</tr>
<tr>
<td>Reinforcement</td>
<td>Messages of congratulations</td>
<td>Notifications</td>
</tr>
<tr>
<td></td>
<td>Weekly emails</td>
<td>Badges</td>
</tr>
<tr>
<td>Communication</td>
<td>Family Forum</td>
<td>Coaching</td>
</tr>
<tr>
<td></td>
<td></td>
<td>State or regional forums</td>
</tr>
<tr>
<td>Message presentation</td>
<td>Simple language</td>
<td>Closed captioning</td>
</tr>
<tr>
<td></td>
<td>Positive and nonstigmatizing tone</td>
<td>Spanish language</td>
</tr>
<tr>
<td></td>
<td>Videos</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pictures</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Font sizes and colors to highlight information</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Checks for understanding in games/quizzes</td>
<td></td>
</tr>
<tr>
<td>Credibility</td>
<td>Evidence-based information from credible source</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Encrypted and password-protected</td>
<td></td>
</tr>
</tbody>
</table>

_a_None provided by participants.

_BT2_: second round of beta-testing.

**Objective 2: Using Qualitative Methods, Evaluate the User Experience of FMF Connect to Guide Further App Refinements**

Next, we considered how users evaluated the FMF Connect intervention and what could be improved to enhance user experience before larger-scale testing. Despite differences in technological problems and usage metrics, iOS and Android users had remarkably similar evaluations of the app. Globally, users were very positive about the app, with 2.5 times more positive- than negative-coded segments across participants. Positive evaluations emphasized the need for and practical utility of the app, which often related to significant barriers to care in this population. The informational and video content of the app was described as particularly valuable to caregivers. The Learning Modules and videos yielded the most detailed discussion, and every caregiver made at least one positive evaluation of the content within the app, much of which was originally derived from the standard FMF Program.

Participants also spoke about the affective and social benefits of the app. They described that raising a child with FASD is a very confusing, frustrating, and isolating experience, which is consistent with previous literature [11,38,39]. Many positive evaluation themes were associated with caregivers’ expressed values of understanding their child and working to support their children’s success. Participants also valued watching videos of real families and connecting with others in the app, which made them feel validated and less isolated. Caregivers’ stated motivators for using the FMF Connect app consistently related to informational content and connecting with others, which aligns well with their values and beliefs.

Negative evaluations of the user experience largely emphasized the technical and navigational aspects of the app. On the basis of feedback from BT1 users, we redesigned the Family Forum interface and added organizational features (eg, table of contents) and tailored video presentation in the Learning Modules to improve navigation. These refinements were positively evaluated by most BT2 users. Consistent with our previous findings during the initial design process [26], the theme of step-by-step progression of learning content received mixed evaluations. All providers and most parents provided at least one positive evaluation of the step-by-step progression of content. Importantly, however, some users (especially more experienced caregivers in usage tiers 2-4) found this progression redundant or tedious. They preferred greater autonomy in self-directed learning, which characterizes the app. Step-by-step progression was identified as a barrier to the use of some caregivers. A recommendation for open access to content occurred frequently in relation to this theme. Time and attentional resources were common barriers to app use, as described by caregivers. Providers also offered insight into
additional barriers that could impact caregiver use of the FMF Connect on the basis of their experiences serving diverse families, including literacy level, English as a second language, caregiver age, and comfort with technology.

Participants offered a large number of valuable recommendations for further app refinement to continue to improve the user experience. As described above, several of these were implemented in the period between BT1 and BT2 and were then favorably evaluated. Additional refinements, such as a behavior tracking tool, changes to weekly emails, and open access to content, have already been implemented and are being tested in a larger feasibility trial. Subsequent refinements, such as a more robust notification system, coaching infrastructure, search tools, overlays, integration of accessibility tools, and optimization of content for viewing and sharing are in progress.

**Relevance of Study Findings for Other Digital or mHealth Parenting Interventions**

Only one other published study has systematically developed and elicited stakeholder feedback on a digital parenting intervention for FASD [40]. In an initial usability study of the Strongest Families intervention, which involved 11 web-based modules and weekly telephone calls, 8 caregivers and 10 providers provided feedback on the intervention across two cycles. Similar to FMF Connect, participants rated the Strongest Families website as appealing and relatively easy to use. Several usability issues were identified, including navigation, amount of content per page, and tailoring of content; these were subsequently refined with generally positive feedback. Together, both studies document the acceptability and feasibility of digital and mHealth interventions for caregivers raising children with FASD. RCTs are underway (Strongest Families; [41]) or planned later this year (FMF Connect). This systematic approach (Figure 1) may serve as a relevant model for the development of other digital and mHealth interventions.

This study demonstrates the benefits of considering the context of parenting values, beliefs, and attitudes when analyzing user evaluations of the FMF Connect app. Indeed, considering the relevant values, beliefs, and attitudes of caregivers will be informative when developing interventions for other clinical populations. Many values expressed by caregivers in this study have been reported by other parents of children with developmental disabilities [42-44]. Themes of needing to educate others and valuing people who understand are common in the developmental disabilities literature [42]. Current findings emphasize how much parents value access to information, especially because they report that many professionals cannot support them. Research suggests that access to information and services is very important for the well-being of parents of children with disabilities, with peers often being the preferred source of information [43,44]. One study found that parents of children with developmental disabilities felt judged and isolated, and often needed to educate others and seek out their own information. These experiences are major stressors for parents [45]. Although these parenting values are reflected in the broader disabilities literature, they may be especially the case in the field of FASD. Research has clearly shown that many professionals lack knowledge and training on FASD [7,46,47]. FASD can also carry stigma, which can lead to increased feelings of judgment and isolation [48,49].

With these points in mind, it is surprising that few digital interventions exist for parents of children with disabilities. Some preliminary evidence shows that website-delivered interventions are effective for parents of children with autism [50], but a significant need for evidence-based, accessible interventions remains. The accessibility of mHealth and digital interventions is responsive to some barriers to care and to the lived experience of parents raising children with disabilities [51]. Motivators of accessing information and connecting with other parents who understand, identified in this study, are likely to generalize to other populations, especially for those with low community awareness and limited access to care. The current findings demonstrate that choice and autonomy are also highly valued for self-directed learning, which is an important consideration for intervention design.

**Strengths and Limitations**

This study represents the first initial test of a mHealth intervention for caregivers raising children with FASD, a part of a systematic approach to app development and evaluation. This study had many strengths, and efforts were made to reduce the impact of the limitations of this study. The findings emphasize the acceptability and feasibility of the FMF Connect app for caregivers and offer important directions for further refinement. This intervention has clear potential for larger-scale dissemination, with vital public health implications for this underserved population—and, perhaps, especially for some subgroups within this population facing greater social disparities. The methodological approach is also rigorous, involving iterative feedback from key stakeholders to ensure relevance and usability, which is an important step in user-centered design and development [20,52].

Study findings are limited by the perspectives of participants sampled. As is true in many studies, all participants were volunteers, contributing to the possibility of selection bias. The current sample size is relatively large for beta-testing feasibility studies and is considered sufficient for the primary objectives of this study. However, it is possible that valuable perspectives may have been missed. Although the consistency of themes was very high across both caregivers and providers, some demographics of this diverse clinical population are less well represented. For example, only 7% of the parents and 6% of the providers were men. Overrepresentation of women is common in studies involving primary caregivers [53,54]. The sample represented primarily adoptive parents, although the perspectives of relative caregivers and biological mothers are represented. Racial and ethnic diversity is also somewhat lower than in the general population, and all participants were English-speaking (although some may have had fluency in other languages). The study was also limited to participants who were able to afford smartphones, WiFi, or data plans. Participant income spanned all queried levels but was still skewed above average relative to national statistics. It is notable that the inclusion of provider perspectives provided additional insights into potential barriers.
and recommendations for families that may not have been represented in the sample.

Although our nonuse attrition rate was generally within the range seen in community and digital interventions, it may have contributed to bias in the study findings. Participants were less likely to complete focus group or individual interviews if they did not install the app or had low use. Therefore, the study of attrition factors was necessarily limited. Interviewees may have displayed a positive response bias. A number of participants expressed desperation for information and resources on FASD; therefore, fewer negative evaluations of the app may have been offered because of the lack of alternative treatments. The research team was also involved in developing the app, which could have impacted participant feedback. To reduce positive response bias, the research team actively tried to elicit negative and constructive feedback about the app and emphasized the benefits of hearing negative evaluations during this stage of development when refinements could more easily be made. Although significantly fewer than positive evaluations (n=788 segments), a robust number of negative evaluations were elicited (n=312 segments). However, it remains a possibility that users experienced difficulty expressing negative feedback in interviews, and this should be considered a limitation of the current work.

Conclusions and Future Directions

This study demonstrates that a scalable digital health intervention can successfully be derived from an empirically supported therapist-led intervention tailored for families raising children with FASD, adding unique and important additional features. The FMF Connect app is acceptable and feasible as self-administered learning for caregivers raising children with FASD. In addition, the FMF Connect app aligns with important reported caregiver values and builds on our previous work evaluating the initial design and functionalities of the app [26]. The sixth phase in our systematic evaluation of the FMF Connect app (Figure 1) is to conduct a larger pilot trial with pre-post quantitative data collection, which is currently ongoing. The results of this pilot trial will provide direction for further refinements to the FMF Connect intervention, measurement tools, and study design methods before the advent of a large-scale RCT. Surprisingly, many mHealth and digital health interventions have been disseminated without clear empirical validation. In our systematic development and evaluation plan, a carefully designed RCT is an important seventh and final phase. This systematic approach is squarely aimed at producing the FMF Connect app as a robust mHealth intervention responsive to the needs of a clinical population that deserves high-quality FASD-informed care.

Acknowledgments

This work was conducted in conjunction with the Collaborative Initiative on Fetal Alcohol Spectrum Disorders, which is funded by grants from the National Institute on Alcohol Abuse and Alcoholism. Additional information about Collaborative Initiative on Fetal Alcohol Spectrum Disorders can be found on the website [55]. This research was supported by the National Institute of Health under Award #U01AA026104. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors’ Contributions

All authors contributed to the preparation of this study. CLMP conceived of the study, led app content adaptation and development, obtained necessary approvals, conducted focus groups and individual interviews, and led data analysis and manuscript preparation. ARR, CCKT, and JEP assisted with app content development, facilitated participant recruitment, conducted interviews, managed data quality, processing, and analysis, and assisted with manuscript preparation. CT led the technological development of the app, assisted with focus group moderation, and contributed to manuscript revisions. UD assisted with use data analytics and manuscript revisions. HCO is the developer of the standard Families Moving Forward (FMF) Program, from which the app is derived. She made significant contributions to intellectual property exchange, app content adaptation and development, and assisted with manuscript revisions.

Seattle Children’s Research Institute (SCRI) and the University of Washington are recognized for their contributions to intellectual property from the standard FMF Program. The standard FMF Program was developed, tested, and materials refined by a team led by author HCO based at and sponsored by these institutions (led by SCRI), with funding via multiple grants from the Centers for Disease Control and Prevention.

The authors appreciate all the families who made time in their busy schedules to participate in this study. This research would not have been possible without their valuable insights. The authors also want to acknowledge Elizabeth Hanlon and Emily Bantelman for their work in transcribing focus group and interview data, and Shuo Zhang and Emily Speybroeck for editing assistance on earlier drafts of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Semistructured questioning route for the focus group and individual interviews.

[DOCX File , 16 KB - formative_v5i12e29687_app1.docx ]
References


35. Percentage of mobile apps that have been used only once from 2010 to 2019. Statista. URL: https://www.statista.com/statistics/271628/percentage-of-apps-used-once-in-the-us/ [accessed 2021-11-07]


55. Collaborative Initiative on Fetal Alcohol Spectrum Disorder. URL: https://www.cifasd.org/ [accessed 2021-11-16]

Abbreviations

- BT1: first round of beta-testing
- BT2: second round of beta-testing
- FASD: fetal alcohol spectrum disorders
- FMF: Families Moving Forward
- mHealth: mobile health
- PAE: prenatal alcohol exposure
- RCT: randomized controlled trial

©Christie Lynn McGee Petrenko, Carson Christine Kautz-Turnbull, Alicia Rose Roth, Jennifer Elizabeth Parr, Cristiano Tapparello, Utku Demir, Heather Carmichael Olson. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.12.2021. 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.
Conversational Agent for Healthy Lifestyle Behavior Change: Web-Based Feasibility Study

Dhakshenya Ardhithy Dhinagaran1, BSc(Hons); Thirunavukkarasu Sathish2,3, MBBS, MPH, PhD; AiJia Soong1, BSc, MPH; Yin-Leng Theng4, PhD; James Best4, MD, MBBS; Lorainne Tudor Car1,5, MD, MSc, PhD

1Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore, Singapore
2Population Health Research Institute, McMaster University, Hamilton, ON, Canada
3Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
4Centre for Healthy and Sustainable Cities, Nanyang Technological University, Singapore, Singapore, Singapore
5Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom

Abstract

Background: The rising incidence of chronic diseases is a growing concern, especially in Singapore, which is one of the high-income countries with the highest prevalence of diabetes. Interventions that promote healthy lifestyle behavior changes have been proven to be effective in reducing the progression of prediabetes to diabetes, but their in-person delivery may not be feasible on a large scale. Novel technologies such as conversational agents are a potential alternative for delivering behavioral interventions that promote healthy lifestyle behavior changes to the public.

Objective: The aim of this study is to assess the feasibility and acceptability of using a conversational agent promoting healthy lifestyle behavior changes in the general population in Singapore.

Methods: We performed a web-based, single-arm feasibility study. The participants were recruited through Facebook over 4 weeks. The Facebook Messenger conversational agent was used to deliver the intervention. The conversations focused on diet, exercise, sleep, and stress and aimed to promote healthy lifestyle behavior changes and improve the participants’ knowledge of diabetes. Messages were sent to the participants four times a week (once for each of the 4 topics of focus) for 4 weeks. We assessed the feasibility of recruitment, defined as at least 75% (150/200) of our target sample of 200 participants in 4 weeks, as well as retention, defined as 33% (66/200) of the recruited sample completing the study. We also assessed the participants’ satisfaction with, and usability of, the conversational agent. In addition, we performed baseline and follow-up assessments of quality of life, diabetes knowledge and risk perception, diet, exercise, sleep, and stress.

Results: We recruited 37.5% (75/200) of the target sample size in 1 month. Of the 75 eligible participants, 60 (80%) provided digital informed consent and completed baseline assessments. Of these 60 participants, 56 (93%) followed the study through till completion. Retention was high at 93% (56/60), along with engagement, denoted by 50% (30/60) of the participants communicating with the conversational agent at each interaction. Acceptability, usability, and satisfaction were generally high. Preliminary efficacy of the intervention showed no definitive improvements in health-related behavior.

Conclusions: The delivery of a conversational agent for healthy lifestyle behavior change through Facebook Messenger was feasible and acceptable. We were unable to recruit our planned sample solely using the free options in Facebook. However, participant retention and conversational agent engagement rates were high. Our findings provide important insights to inform the design of a future randomized controlled trial.

(JMIR Form Res 2021;5(12):e27956) doi:10.2196/27956
KEYWORDS
chatbot; conversational agents; behavior change; healthy lifestyle behavior change; pilot study; feasibility trial; usability; acceptability; preliminary efficacy; mobile phone

Introduction

Background
In recent years, there has been a notable increase in the incidence of chronic disease, especially among the younger population [1]. These chronic diseases include obesity and type 2 diabetes [1]. In terms of diabetes prevalence, Singapore, with 600,000 adults living with diabetes, ranks second among high-income countries, and obesity levels are also on the rise [2]. In addition, 34% of the men and 39% of the women in Singapore do not reach the weekly target of 150 minutes of moderate-intensity activity per week, leaving people at higher risk of developing diseases such as diabetes [3]. People with diabetes live shorter lives—by at least 10 years—and have a lower quality of life (QoL) than those without diabetes [4,5]. Prediabetes, a precursor to diabetes, affected 15.5% of the Singaporean adults in 2010 (ie, 1 in 7), and this figure is estimated to increase to 24.9% by 2035 [6]. Prediabetes increases the risk of heart disease, and if untreated, over time, most people with prediabetes transition to diabetes [7]. Research has shown that lifestyle interventions (such as increasing physical activity and eating a healthy diet) delivered by trained health care professionals can help to promote healthy weight loss [8-10]. In addition, lifestyle change through high-risk and population-based approaches were also directly associated with a reduction in the incidence of type 2 diabetes, accentuating the efficacy of healthy lifestyle behavior change for diabetes prevention [8].

Achieving healthy lifestyle behavior changes independently can be challenging, and support from experts such as dietitians or exercise physiologists has been shown to be more effective [8]. However, access to experts at a population level may not be feasible or affordable. A potentially more accessible alternative to in-person support and supervision could be novel digital health interventions such as conversational agents. Conversational agents, or chatbots, are computer programs designed to mimic human-to-human conversations in the form of either text messaging or verbal discourse [11]. The heightened accessibility, personalization, and efficiency that conversational agents offer highlight the potential for conversational agents to improve patient care [11-13]. Conversational agents enable 2-way communication, and their text- or speech-based method of communication makes them suitable for a variety of target populations, ranging from young children to older people. The application of conversational agents in health care is gaining traction in a number of medical fields, including health care service provision, chronic disease management, and patient education [14]. They can be delivered through a variety of means: messaging apps, individual apps, or even standalone devices [14].

Singapore is a technologically savvy country, and citizens avidly use messaging apps. In addition, Singapore’s ministry of health has proposed the increasing use of conversational agents in health care in tackling issues such as the rising chronic disease burden and the aging population [15] that can lead to more primary care appointments. The ministry envisions a near future where a conversational agent can collect a patient’s history from them before their consultation, streamlining the primary care visit and thus cutting down waiting times. This makes Singapore an ideal place for the evaluation of novel mobile health interventions such as conversational agents. Moreover, health programs delivered over the internet have shown success, as exemplified by web-based interventions for smoking [16], alcohol intake [17], sexual health [18], cancer screening [19], physical activity [20], and diet modifications [21]. The ubiquity of the internet makes these programs easily accessible to a diverse group.

Objective
The evidence for the use of conversational agents for healthy lifestyle behavior change from trials is limited. The feasibility and acceptability of implementing and evaluating the use of novel interventions are essential for informing potential future trials. Correspondingly, we aim to assess the feasibility, acceptability, and preliminary efficacy of the use of conversational agents for healthy lifestyle behavior changes in the general population in Singapore.

Methods

Approval and Consent
This web-based single-arm feasibility study was approved by the Nanyang Technological University Ethics Committee (IRB-2018-11-032). All participants signed their digital informed consent before embarking on the study.

Participants
Participants were eligible if they were aged above 21 years, were Singapore citizens or permanent residents, owned a smartphone, and had a Facebook Messenger (Facebook, Inc) account. Prospective participants were excluded if they were pregnant or had any of the following conditions: cancer, chronic liver disease, chronic kidney disease, a neurodegenerative condition, heart disease, stroke, a physical disability, hypertension, or a condition that does not allow for regular physical activity. Eligibility was confirmed by having participants complete an eligibility questionnaire, after which they were asked to provide informed consent on a digital form sent to them through email.

Recruitment
The participants were recruited on the web through Facebook in August-September 2019. A digital poster listing the study aims and eligibility criteria was uploaded on relevant Facebook pages focused on healthy living, such as Singapore fitness and health community and Singapore healthy cooking. In addition, we used snowballing in our recruitment; therefore, participants were also procured through redistribution of our study poster through messaging apps or through word of mouth.
Intervention

The conversational agent was designed to be used on Facebook Messenger using a free web-based tool, Chatfuel [22]. A research associate (AS) and a PhD student (DAD) developed the script for the program and performed the input. The intervention focused on diabetes and prediabetes knowledge, diet, physical activity, sleep, and stress management. These were the topics of focus identified in other diabetes prevention programs targeting lifestyle change [23-25].

The content was informed by existing evidence-based sources of information, including clinical guidelines and systematic reviews. Advice on improving sleep quality was generated from published evidence reporting on techniques and successful interventions for sleep disorders [26]. The domain on stress was informed by distance learning–based stress management techniques identified from a review of existing studies that described methods to reduce stress and improve health [27]. Pertinent nutritional advice for individuals with prediabetes was obtained from authenticated government health portals and other validated health and nutrition webpages [28]. The collated advice was then compartmentalized into themes, which translated into the topic of focus for each interaction between conversational agent and user (Multimedia Appendix 1 [8,26,27,29-53]). The content for the section on physical activity was informed by advice on the recommended duration and intensity of exercise from Singapore’s Health Promotion Board [28]. Validated fitness routines were then presented as part of the conversation as methods to achieve the necessary level of fitness [29]. Examples of conversational exchanges between the conversational agent and users are presented in Figure 1.

Figure 1. A representation of what the interactions between Precilla and users entailed.

We also followed the Capability, Opportunity, Motivation, Behavior model of behavior to guide the development of the intervention [54]. This model posits that to exhibit a particular behavior (B), the participant must be physically and psychologically capable (C), have sufficient social and physical opportunities (O) to perform the behavior, and must have the desire or need to do so—motivation (M). This was important in determining the inclusion criteria for the intervention (eg, physically fit with no comorbidities) and in conversation designing.

The content was mostly presented in the form of text, supplemented with some images to make the conversational agent more engaging and to enhance the user experience. These images were obtained from free-to-use sources or photographs taken by our study team members. To further contribute to a positive user experience that mimicked human interaction, the
A conversational agent was given a name, Precilla, and it displayed human-like characteristics such as in the tone of speech, profile picture, and through using the typing function for messages. From here onward, we use Precilla to refer to the conversational agent used in our pilot study.

The intervention was scheduled to last for 4 weeks, whereby participants would receive 4 messages per week, (every other day). 1 for each topic (diet, exercise, sleep, and stress). A sample of the dialog tree is presented in Multimedia Appendix 2.

Acceptability of the Intervention

The acceptability of Precilla was measured through questions on usability and satisfaction in the follow-up questionnaire. Questions were asked on the participants’ overall satisfaction with Precilla, their likelihood of using Precilla again and recommending it to others, as well as the impact of the interactions on their health.

The usability questionnaire was split into 2 sections; the questions in section 1 related to how participants perceived the usability and their overall satisfaction with the study. Interested participants also had an opportunity to take part in a follow-up interview to share their views on the conversational agent, their experience while taking part in this web-based study, as well as their thoughts on points of improvement.

Outcomes

Primary Outcomes

The primary outcomes of interest for this study were the feasibility of recruitment and retention of participants, acceptability of the intervention, and participant engagement with the intervention.

Feasibility of Recruitment and Retention

Feasibility in this study was determined by recruitment and retention. Feasibility of recruitment was defined as the ability of the researchers to recruit at least 75% (150/200) of the target sample on the web using Facebook within a 1-month period. Feasibility of retaining participants was defined as at least 32.7% (49/150) of the recruited participants completing the study.

Secondary Outcomes

Our secondary outcomes were related to the efficacy of the intervention with regard to the participants’ QoL, diabetes knowledge, diet, physical activity, sleep, and stress over 4 weeks.

QoL was measured using the short form version of the QoL Enjoyment and Satisfaction Questionnaire [56]. The questionnaire comprises 14 items that are rated on a 5-point scale that indicates the degree of enjoyment or satisfaction experienced during the past week. The total score for the 14 items, which cover the topics of work, social life, health, and overall well-being, ranges from 14 to 70. A percentage of the maximum score is also reported; for example, if a participant

Procedure

An overview of the study workflow is demonstrated in Figure 2. Interested individuals were first required to complete an eligibility questionnaire. The eligible participants then provided informed consent, after which they were requested to complete a baseline questionnaire on a web-based tool, REDCap (Research Electronic Data Capture; Vanderbilt University) [55]. The questionnaire obtained general information regarding demographics as well as data on diet, physical activity, sleep, and stress.

Figure 2. Study workflow.
scores 20, the percentage—29% (20/70)—is also reported. The last 2 questions on medication adherence and overall life satisfaction were reported as percentages (Multimedia Appendix 3).

The knowledge questionnaire had 3 separate sections. Section 1 was concerned with participants’ knowledge about healthy living, prediabetes, and diabetes; section 2 asked for their perceptions on how healthy their lifestyle is and the likelihood of their developing diabetes; section 3 tested their knowledge of risk factors for type 2 diabetes. The questionnaire was derived from an adaptation of questions presented by the Michigan Diabetes Research Centre [57]. The questionnaire was adapted such that, in section 1, questions regarding the definitions of diabetes and prediabetes were added to test participants’ knowledge of these conditions. In section 2, the wording of some questions about participants’ efforts to make healthy lifestyle behavior changes in the past year was adapted to obtain more detail in their responses from just yes or no to a scaled format. In section 3, some risk factors that were not applicable to the Singaporean context were removed (eg, being Asian American, Hispanic, or African American).

The participants’ diets were assessed based on an adaptation of the Food Frequency Questionnaire where questions were asked on the frequency of daily intake of vegetables, fruits, fried foods, and sweetened drinks [58]. Physical activity was assessed using the International Physical Activity Questionnaire (IPAQ), which was adapted specifically for use in this study [59]. Questions were asked on the intensity of exercise (vigorous, moderate, or light), the frequency of physical activity (in days per week), and the length of each session (in minutes). Only sessions lasting at least 10 minutes qualified as physical activity. In line with IPAQ scoring, a metabolic equivalent of task (MET) score was calculated. The MET score represented the amount of energy expended when carrying out physical activity. For consistency, walking was given a score of 3.3 METs, moderate physical activity 4 METs, and vigorous physical activity 8 METs. To calculate MET minutes per week, the MET value was multiplied by the minutes for which the activity was carried out and again by the number of days in the week that the activity was undertaken. As some participants provided a range for their responses, for consistency, a mean value was used for the calculation. For example, for number of days, that is, 1-3 days, the mean was calculated as 2. For session length, that is, 10-20 minutes, the mean was calculated as 15.

A MET score of 600 METs per week indicated that an individual was moderately physically active, whereas 1500 METs per week indicated a high level of physical activity. Any score that did not qualify as moderate or high was considered an indication of a low level of physical activity.

Sleep scoring was done using the Pittsburgh Sleep Quality Index (PSQI) questionnaire [60]. A score of 0 indicated excellent sleep quality, whereas 10 indicated severely poor sleep quality. Poor sleep quality was defined as participants with a global PSQI score higher than 5. The stress level of the participants was gauged using the Perceived Stress Scale, where it was possible to receive a score between 0 and 40 [61]. Scores ranging from 0 to 13 were considered low perceived stress, from 14 to 26 moderate perceived stress, and from 27 to 40 high perceived stress.

### Data Analysis

Data analysis was conducted on the data collected at baseline and follow-up (Table 1). The participant responses on satisfaction and usability were presented as percentages of each option on a Likert scale (eg, strongly agree, disagree, much better than before, or neither better nor worse). The participant outcomes for QoL, knowledge, physical activity, sleep, and stress were all presented as pre- and postscores using the individualized scoring system of each questionnaire.

**Table 1.** Overall summary of results (N=60).

<table>
<thead>
<tr>
<th>Knowledge, mean (SD; range)</th>
<th>Baseline values (n=60)</th>
<th>Follow-up (4 weeks) values (n=56)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 1</td>
<td>7 (1.37; 6-18)</td>
<td>6 (0.93; 6-18)</td>
</tr>
<tr>
<td>Section 2</td>
<td>28 (2.7; 12-48)</td>
<td>27 (3.5; 12-48)</td>
</tr>
<tr>
<td>Section 3</td>
<td>10 (4.1; 8-32)</td>
<td>9 (2.2; 8-32)</td>
</tr>
<tr>
<td>Sleep, mean (SD)</td>
<td>4 (2.36)</td>
<td>4 (2.45)</td>
</tr>
<tr>
<td>Stress, mean (SD)</td>
<td>17 (5.13)</td>
<td>16 (5.10)</td>
</tr>
<tr>
<td>Physical activity (MET(^a) score), mean (SD)</td>
<td>1080 (816)</td>
<td>1075 (872)</td>
</tr>
<tr>
<td>QoL(^b)</td>
<td>54 (6.90)</td>
<td>53 (6.46)</td>
</tr>
<tr>
<td>Maximum score (%)</td>
<td>77</td>
<td>75</td>
</tr>
</tbody>
</table>

\(^a\)MET: metabolic equivalent of task.

\(^b\)QoL: quality of life (measured using the short form version of the QoL Enjoyment and Satisfaction Questionnaire).

We performed descriptive analyses of the data. The data were presented using percentages, means, and SDs. As this was a feasibility study, no hypothesis testing was performed to assess the efficacy of the intervention [62]. Mean differences between baseline and follow-up were presented, accompanied by 95% CIs [62].
Results

Recruitment and Retention
A total of 136 individuals expressed initial interest in participating in the study; however, 9 (6.6%) were ineligible because of comorbidities or because they had not installed Facebook Messenger on their smartphone, and a further 52 (38.2%) completed the screening questionnaire but did not proceed to provide informed consent, leaving 75 (55.1%) participants eligible for participation. Of these 75 participants, 60 (80%) completed baseline assessments, and of these 60, 56 (93%) completed the follow-up questionnaires (Figure 3).

Figure 3. The flow of participants through the study.

Characteristics of Participants
Of the enrolled participants, 38% (23/60) were men. The mean age was 33.7 years (SD 9.3), the mean BMI was 22.3 kg/m² (SD 3.8; Table 2), and the average technical competency on a scale of 1 to 10 was 8.07.
Table 2. Characteristics of all the enrolled participants who completed baseline assessments (N=60).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Baseline values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>33.7 (9.3)</td>
</tr>
<tr>
<td>BMI (kg/m²), mean (SD)</td>
<td>22.3 (3.8)</td>
</tr>
<tr>
<td>Gender (male), n (%)</td>
<td>23 (38)</td>
</tr>
</tbody>
</table>

**Ethnicity, n (%)**
- Chinese                                  | 48 (80)         |
- Malay                                    | 5 (8)           |
- Indian                                   | 3 (5)           |
- Others                                   |                 |
  - White                                   | 2 (3)           |
  - Burmese                                 | 2 (3)           |

**Marital status, n (%)**
- Currently married                        | 25 (42)         |
- Never married                            | 33 (55)         |
- Separated                                | 1 (2)           |
- Divorced                                 | 1 (2)           |

**Highest level of education, n (%)**
- University and above                     | 50 (83)         |
- Polytechnic diploma                      | 1 (2)           |
- Other diploma and professional qualification | 1 (2)   |
- A²-level or NTC⁻¹ or NTC-2 or certificate in office or business skills or its equivalent | 6 (10) |
- O³ or N⁵-level or NTC-3 certificate or its equivalent | 1 (2) |
- Secondary school                         | 1 (2)           |

**Work status, n (%)**
- Employed                                 | 37 (62)         |
- Student (full time)                      | 17 (28)         |
- Homemaker or housewife                   | 3 (5)           |
- Unemployed (able to work)                | 2 (3)           |
- Retired                                  | 1 (2)           |

**History of parents, sibling, or child with type 2 diabetes, n (%)**
- Yes                                      | 11 (18)         |
- No                                       | 49 (82)         |

**History of hypertension, n (%)**
- Yes                                      | 2 (3)           |
- No                                       | 58 (97)         |

---

a A-level or advanced-level examinations are taken by students at the age of 18, 2 or 3 years after completing their O-level or ordinary-level examinations, which are taken by students at the age of 16 after 4 years (or 5 years) of secondary school.

b NTC: National Technical Certificate.

N-level: The Singapore-Cambridge General Certificate of Education Normal-level. Secondary students in Singapore can move between two streams based on their academic performance: 4 years of study culminating in the O-level (ordinary level) or the N-level (normal level) examinations. N-level students may participate in a fifth year of study to take the O-level examinations.
Satisfaction

Of the 56 participants who followed the study through till completion, 52 (92%) were moderately or very satisfied with Precilla, 30 (54%) thought that their likelihood of recommending Precilla to others would be somewhat likely, 32 (57%) felt that the likelihood of their opting to use Precilla again for personal use would also be somewhat likely, and, finally, 29 (51%) posited that their health was somewhat better or much better than before.

Usability

The agree and strongly agree responses exceeded 50% (28/56) for all the questions relating to Precilla’s acceptability, except for question 5, “Chatting with Precilla motivated me to change unhealthy sleeping habits,” where the collective response in agreement (agree and strongly agree) was only 46% (26/56) (Figure 4). The smallest percentage of disagreement was for questions 1 and 8, indicating that most of the participants (55/56, 98%) found the chat easy to use and thought that Precilla used simple language and was easy to talk to. The neutral response was a common selection by participants for questions 2-7.

Figure 4. Participants’ opinions of Precilla’s usability and acceptability.

Of the 56 participants, 37 (67%) agreed that Facebook Messenger was an appropriate medium to deliver the messages sent by Precilla. The remaining participants were either neutral or disagreed. Telegram and WhatsApp were proposed as potential alternative delivery media. Of the 56 participants, 9 (15%) thought that the number of messages exchanged with Precilla was too high, and they would have preferred the intervention to deliver fewer messages; 12 (21%) disagreed that 12 PM was the most appropriate time to receive messages, whereas the remaining 44 (79%) either agreed, strongly agreed, or were neutral (Figure 4) and indicated a preference for interacting with Precilla after work, in the evening or before bedtime. Some indicated specific timings such as 5 PM, 8 PM, or 10 PM, and 1 participant suggested 9 AM or during the morning commute to work. There were no disagreements to statement 13 (“I liked images and videos in Precilla’s messages”) and minimal disagreement for statements 14 (5/56, 9%) and 15 (4/56, 7%) on the inclusion of more visual components and Precilla’s personality, respectively. Most participants did not find Precilla to be very human-like and could clearly tell that they were communicating with a conversational agent. Of the 56 participants, 3 (5%) suggested some degree of personalization, whereby Precilla’s messages should be initiated when the user prompted the conversational agent and not the other way around.

Conversational Agent Engagement Data

The number of complete, incomplete, and absent interactions were noted (Figure 5). In addition, the number of immediate and delayed interactions were also counted (Figure 6). Some interactions were started but not completed. The reasons for not completing interactions are presented in Multimedia Appendix 4.
All the conversations were completed by at least 50% (>30/56) of the participants consistently over the period of the study. The number of participants who completed interactions from start to finish exceeded 71% (40/56) for 56% (9/16) of the interactions, denoting high engagement with the conversational agent. Similarly, at least 40% (<22/56) of the participants responded immediately to the conversational agent 75% (12/16 interactions) of the time, even in the last week of the study, denoting consistently high conversational agent engagement.

The participants were restricted to a predetermined list of options when providing responses to the questions asked or information provided. A frequent cause of disruption to the conversation flow was participants typing free text, which led them off track. This resulted in their being redirected to the main menu, which they potentially found confusing, or they were reluctant to go through the whole conversation flow again and, hence, failed to complete that interaction. The average duration of each interaction was in the 2- to 5-minute range. More data on conversational agent engagement are presented in Multimedia Appendix 4.

**QoL Score**

The 14-item QoL score was 54 (SD 6.90) at baseline and 53 (SD 6.46) at follow-up (mean difference 0.95; 95% CI −1.51 to 3.41). Of the participants taking medication (10 at baseline, 17 at follow up), participants rated adherence as fair (baseline
Discussion

Principal Findings

To our knowledge, this is the first web-based pilot study to test the feasibility and acceptability of a healthy lifestyle behavior change conversational agent in the Singaporean population. In this web-based study, which used Facebook Messenger as a delivery medium for a conversational agent targeting healthy lifestyle behavior change in the Singaporean population, we managed to recruit 37.5% (75/200) of the target sample size in 1 month. Retention was high at 93% (56/60); conversational agent engagement was also high, with all the conversations being completed by at least 50% (28/56) of the participants consistently. Furthermore, at least 40% (22/56) of the participants responded almost immediately 75% (12/16) of the time. Acceptability, usability, and satisfaction were also generally high. In general, we were able to conduct the study with high fidelity, and each phase ensued as planned. Any definitive signs of improvement in QoL, knowledge, diet, physical activity, sleep, and stress would have to be studied in more detail in a future study of effectiveness.

The secondary outcomes in the study were measured using validated scales, namely, the short form version of the QoL Enjoyment and Satisfaction Questionnaire, PSQI, Perceived Stress Scale, Food Frequency Questionnaire, and IPAQ. We took note that although these questionnaires have not necessarily been optimized for the Singaporean population, these well-established scales were chosen to ensure a certain level of validity and reliability in our results.

Comparisons With Existing Literature

Conversational agents have been used for healthy behavior change; however, they have been used mainly in niche areas such as smoking cessation, alcohol misuse treatment, and the promotion of physical activity in sedentary populations [63-65]. The application presented in this study is very comprehensive, novel, and relevant to the general population in Singapore.

The outcome measures were similar to those in other studies looking at the acceptability of conversational agents in health care, and acceptability was high in other studies as well as in this feasibility trial. Previous studies noted high response rates (ie, conversational agent engagement) and strongly agree or agree scores for user-friendliness, appropriateness, consistency, and speed of response, as in our study [66]. The other measures in prior studies that denoted acceptability were high perceived ease of use, usefulness, and intention to use, similar to our measures of ease and enjoyment of use, motivating change in unhealthy habits, and intention to use Precilla again or recommend it to others, respectively [67,68].

In another study using a health behavior change conversational agent, high compliance was attributed to a rewarding game component, which was consistent with high fidelity, and each phase ensued as planned. Any definitive signs of improvement in QoL, knowledge, diet, physical activity, sleep, and stress would have to be studied in more detail in a future study of effectiveness.

In summary, the study results will add to the growing body of evidence regarding the acceptability of conversational agents for healthy behavior change in Singapore, and the development of conversational agents targeting lifestyle behavior change in Singapore is feasible.

5/10, 50%; follow-up 6/17, 35%), good (baseline 5/10, 50%; follow-up 9/17, 53%), and very good (baseline 0/10, 0%; follow-up 2/17, 12%). The participants rated overall life satisfaction as fair (baseline 16/60, 27%; follow-up 12/56, 21%), good (baseline 35/60, 58%; follow-up 33/56, 59%), or very good (baseline 9/60, 15%; follow-up 10/56, 18%), and 2% (1/56) of participants reported poor overall life satisfaction at follow-up.

Knowledge

The scoring system for the questionnaires meant that lower scores denoted better knowledge outcomes (sections 1 and 3) and lower perceived risk of diabetes (section 2). In section 1, the scores were 7 (SD 1.37) at baseline and 6 (SD 0.93) at follow-up (mean difference 0.52, 95% CI 0.09-0.95); the possible score range was 6-18. In section 2, the scores were 28 (SD 2.7) at baseline and 27 (SD 3.5) at follow-up (mean difference 0.9, 95% CI −0.25 to 2.05); the possible score range was 12-48. In section 3, the scores were 10 (SD 4.11) at baseline and 9 (SD 2.2) at follow-up (mean difference 1.0, 95% CI −0.23 to 2.23); the possible score range was 8-32.

Diet

The number of individuals consuming vegetables at least once a day was 27% (16/60) at baseline and 29% (16/56) at follow-up. Similarly, the percentage of participants having three portions of fruit per day was 3% (2/60) at baseline and 7% (4/56) at follow-up. With regard to intake of unhealthy food comprising sweetened beverages and fried food or snacks, the almost never consumption category for sweetened beverages was 38% (23/60) at baseline and 45% (25/56) at follow-up, whereas for fried food and snacks, the almost never consumption category was 25% (15/60) at baseline and 30% (17/56) at follow-up.

Physical Activity

The median METs-per-week score was 857 at baseline (IQR 902) and 765 (IQR 882) at follow-up. The average categorical score remained constant from baseline to follow-up where the average exercise intensity was moderate. The mean time spent sitting was 439 minutes at baseline and 406 minutes at follow-up. In addition, the median amount of moderate and vigorous physical activity was the same, 30 min per week (IQR 90) at baseline and 50 min per week (IQR 90) at follow-up.

Sleep

According to the scoring method, a lower score indicated better sleep. The mean PSQI score was 4.38 (SD 2.36) at baseline and 4.43 (SD 2.45) at follow-up (mean difference −0.05, 95% CI −0.93 to 0.83). The prevalence of poor sleep quality was 28% (16/60) at baseline and 30% (16/56) at follow-up. The participants who reported sleeping for less than 7 hours were classified as short sleepers; 82% (49/60) were short sleepers at baseline, whereas 88% (49/56) were short sleepers at follow-up.

Stress

The scoring for stress meant that lower scores denoted lower stress levels and higher scores denoted higher stress levels. The mean score at baseline was 17 (SD 5.1), denoting a moderate stress level, which was maintained even at follow-up, mean 16 (SD 5.1; mean difference 0.73; 95% CI −1.15 to 2.61).
by providing personalized advice to maintain target blood glucose levels or personalized reminders for taking medication [69]. Similarly, personalized content and message timing delivery should be implemented in future versions of Precilla, considering the preference indicated by users.

It has previously been indicated that users tend to prefer interacting with female conversational agents, as evidenced by Siri, Cortana, and Alexa, for example [70]. In addition, the study by Brahnam et al [71] has explained that in the field of human-computer interaction, “the standard of believability has become inextricably linked to gender personification, especially female personification.” As such, we chose to use Precilla and reinforced her character with a profile picture.

Views on conversational agent personality have been mixed, depending on the function of the agent. For example, participants appreciated an empathetic demeanor from conversational agents for e-therapy, in contrast to the participants’ indifference to Precilla’s humanity in this pilot study, where she played more of an impersonal, informative or educational role [72]. It may be that the conversational agent’s purpose determines how important the degree of humanity is.

Another aspect of this study was the effectiveness of participant recruitment through social media. We chose a goal of 75% (150/200) recruitment rate based on other studies involving mobile health trials having achieved an application rate of 86% (70/81) for enrollment through Facebook or an 85.1% (605/711) response rate to Facebook advertisements in 45 days [73,74]. Although our study only yielded a recruitment rate of 37.5% (75/200) in the given period, it aligns quite appropriately with the recruitment rates of 33%, 30%, or 37.7% in other web-based mobile apps or Facebook trials [75-77]. Hence, our 37.5% (75/200) recruitment rate seems to indicate feasibility regarding the recruitment of participants in the general population in Singapore for a web-based conversational agent intervention delivered through Facebook.

A systematic review accentuated the effectiveness of Facebook as a recruitment tool [78]. It was suggested that on average, studies allowed a 3-month recruitment period, and Facebook was found to be more efficient than traditional methods (print, radio, television, email, or word of mouth) because of the reduced costs, shorter recruitment times, and ability to connect with harder-to-reach populations. In future studies, we may need to consider complementing our current no-fee Facebook approach with advertisements or other recruitment methods, as well as potentially a longer recruitment period.

**Implications and Future Research**

In terms of the recruitment, we noticed some attrition among the eligible participants just before provision of informed consent. As the only communication was through email from an institutional email address, the uncertainty involved in remote participation, such as never meeting the study team in person, could have been a cause for the participants’ apprehension. This lack of direct engagement may have led to some participants not feeling comfortable sharing their digital signature with us because of data protection and privacy concerns. In future research studies, a mix of a digital recruitment approach and face-to-face exchange (possibly in the form of recruitment or debriefing) could provide participants with an opportunity to ask questions and validate the legitimacy of the study.

Regarding the preliminary efficacy, minimal improvements were observed in the participants’ knowledge, stress, and diet, whereas there was a lack of improvement in QoL and sleep and no change in physical activity. The dearth of significant improvement in these areas could be attributed to the content and delivery not being adequately designed to target effectiveness and, possibly, the short study length (4 weeks), which may not have been sufficient for any noteworthy changes to have been observed. Furthermore, this study was not designed to test the effectiveness of the intervention, and this should be explored in more detail in future iterations.

The participants indicated that they could clearly tell that Precilla was indeed a conversational agent. This was beneficial in confirming Precilla’s identity and in reducing expectations from users that the conversational agent should think, function, and react like a human being. It was interesting to note that although the participants did not find Precilla to be very human-like, they still rated the conversational agent highly in terms of content, usability, general acceptability, ease of use, visual components, and so on. This shows that, for this study population and for a conversational agent with Precilla’s functions and capabilities, acceptability was not dependent on the anthropomorphism of the conversational agent.

The conversational agent engagement data revealed that there was a slight decrease in the number of interactions that the participants completed fully with each successive week. It is possible that the content and delivery methods could have come across as repetitive because we provided a review of take-home points from the previous interaction from week 2 onwards. Alternatively, perhaps the routine that Precilla followed made the intervention very regimented, not leaving much room for spontaneity in the form of varied message timings or even variations in the ways that users could respond to Precilla, for example, using free text. These factors may have contributed to the decrease in interest over time. Future iterations of Precilla could explore introducing more novelty and personalization, as recommended by the participants in the telephone interviews. The participants shared some points for improvement, including alternative modes of delivery such as WhatsApp and Telegram, a possibility of shortening the conversation lengths, and more personalized timings for message delivery.

**Strengths and Limitations**

One of the main strengths of the study is the high fidelity in the delivery of this low-cost, fully web-based feasibility study. We did not experience any software malfunctions and were able to implement the intervention as per protocol. The content of the intervention was evidence based, and it was co-designed with members of the target population to make the intervention as relevant to them as possible. We managed to reach a diverse range of age groups—from 23 to 60 years—using Facebook without the need to pay for advertising. Although the recruited sample—recruited over 4 weeks using solely Facebook—consisted of fewer participants than our target size, it is comparable with those reported in other studies that used...
paid Facebook advertisements within the same time frame. However, the participants were recruited through healthy living–focused Facebook groups and pages; therefore, these individuals were already very much invested in healthy lifestyle behavior change and more likely to use apps and social media. Furthermore, the participants were Facebook users, capable of navigating Facebook Messenger, had high technological competency, and a very high level of tertiary education. As such, our findings may not be generalizable to other demographic groups such as older people; individuals with lower education levels; or those not active on, or familiar with, social media. However, this study did largely cover the general characteristics of the average Singaporean, making the findings very much valuable and relevant. A further limitation is that although we used validated outcome measurement tools, all outcomes were self-reported by the participants. Furthermore, although these measurement scales are well reputed and established, they were not necessarily optimized for use in the Singaporean population. Finally, we used a single-arm pre-post study design, which is acceptable for a feasibility study but does not allow for assessment of the effectiveness of the intervention.

Conclusions
This web-based feasibility study showed that the delivery of a conversational agent for healthy lifestyle behavior change using Facebook Messenger is, to a large extent, feasible in Singapore. Precilla is a low-cost intervention that was popular among the participants and was well received, with most participants being satisfied with the intervention and prepared to recommend it to friends and family. This study demonstrated the ability to conduct a web-based trial to assess the impact of a novel intervention. Our preliminary data on the acceptability of the intervention showed the need for further enhancement of this conversational agent intervention, potentially through humanization of the agent and personalization of the messaging. Such an intervention needs to be evaluated with a rigorous study design and larger sample size.

Acknowledgments
This research was supported by the Ageing Research Institute for Society and Education, Nanyang Technological University, Singapore. This study was also supported by the National Research Foundation, Prime Minister’s Office, Singapore, under its Campus for Research Excellence and Technological Enterprise program.

Authors’ Contributions
LTC conceived the idea for this study. DAD, AS, and LTC developed the resources and administered the study. DAD recruited the participants and collected the data. DAD analyzed the data and interpreted the results with input from TS and LTC. DAD, TS, LTC, AS, JB, and YLT were involved in writing the paper and approved the submitted version.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Sources of information for the conversational agent.
[DOCX File, 19 KB - formative_v5i12e27956_app1.docx ]

Multimedia Appendix 2
Dialogue tree sample outline for introductory week. The message blocks are in grey, and answer button options in blue. User’s name will be seen in the {first name} area. The “typing” lasts between 2-5 seconds depending on the length of text to be read.
[DOCX File, 86 KB - formative_v5i12e27956_app2.docx ]

Multimedia Appendix 3
Questionnaires administered to participants.
[DOCX File, 26 KB - formative_v5i12e27956_app3.docx ]

Multimedia Appendix 4
Conversational agent engagement data.
[DOCX File, 18 KB - formative_v5i12e27956_app4.docx ]

References


28. Aim for 150 minutes of physical activity every week. Health hub. URL: https://www.healthhub.sg/live-healthy/386/150minutes [accessed 2021-09-20]


43. Too early to get up, too late to get back to sleep. Harvard Health Publishing. 2020. URL: https://www.health.harvard.edu/staying-healthy/too-early-to-get-up-too-late-to-get-back-to-sleep [accessed 2021-10-04]


58. Food frequency questionnaires. NIHR Cambridge Biomedical Research Centre. URL: https://dapa-toolkit.mrc.ac.uk/diet/questionnaires/food-frequency-questionnaire [accessed 2021-09-20]


Abbreviations

- **IPAQ:** International Physical Activity Questionnaire
- **MET:** metabolic equivalent of task
- **PSQI:** Pittsburgh Sleep Quality Index
- **QoL:** quality of life
- **REDCap:** Research Electronic Data Capture

©Dhakshenya Ardhithy Dhinagaran, Thirunavukkarasu Sathish, AiJia Soong, Yin-Leng Theng, James Best, Lorainne Tudor Car. Originally published in JMIR Formative Research (https://formative.jmir.org), 03.12.2021. 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.
Patient-Reported Outcome Measure for Real-time Symptom Assessment in Women With Endometriosis: Focus Group Study

Esther van Barneveld¹, MD; Arianne Lim¹,², MD, PhD; Nehalennia van Hanegem³, MD, PhD; Lisa Vork⁴,⁵, MD; Alexandra Herrewegh⁶,⁷, MD; Mikal van Poll⁸, MD; Jessica Manders⁸, MD; Frits van Osch⁵,⁹, PhD; Wilbert Spaans¹,², MD; Gommer van Koeveringe⁶,⁷, MD, PhD; Desiree Vrijens⁶,⁷, MD, PhD; Joanna Kruimel⁴,⁵, MD, PhD; Marlies Bongers¹,²,¹⁰, MD, PhD; Carsten Leue⁷,¹¹, MD, PhD

¹Department of Gynaecology and Obstetrics, Maastricht University Medical Centre+, Maastricht, Netherlands
²Grow School of Oncology and Developmental Biology, Maastricht University, Maastricht, Netherlands
³Department of Gynaecology and Obstetrics, University Medical Centre Utrecht, Utrecht, Netherlands
⁴Department of Internal Medicine, Division of Gastroenterology and Hepatology, Maastricht University Medical Centre+, Maastricht, Netherlands
⁵NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University, Maastricht, Netherlands
⁶Department of Urology, Maastricht University Medical Centre+, Maastricht, Netherlands
⁷School for Mental Health and Neuroscience (MeHNS), Maastricht University, Maastricht, Netherlands
⁸Maastricht University, Maastricht, Netherlands
⁹Department of Clinical Epidemiology, VieCuri Medical Centre, Venlo, Netherlands
¹⁰Department of Gynaecology and Obstetrics, Máxima Medical Centre, Veldhoven, Netherlands
¹¹Department of Psychiatry and Psychology, Maastricht University Medical Centre+, Maastricht, Netherlands

Corresponding Author:
Esther van Barneveld, MD
Department of Gynaecology and Obstetrics
Maastricht University Medical Centre+
P. Debyeelaan 25
Maastricht, 6229 HX
Netherlands
Phone: 31 433876543
Email: esthervb@gmail.com

Abstract

Background: Symptoms related to endometriosis have a significant impact on the quality of life, and symptoms often recur. The experience sampling method (ESM), a digital questioning method characterized by randomly repeated momentary assessments, has several advantages over traditionally used measurements, including the ability to assess the temporal relationship between variables such as physical, mental, and social factors.

Objective: The aim of this study is to develop an ESM tool for patients with endometriosis to accurately measure symptoms and their course over time, allowing for personalized treatment and adequate monitoring of treatment efficacy in individual patients.

Methods: On the basis of international guidelines, items from validated questionnaires were selected through a literature review and during focus groups and multidisciplinary expert meetings. Data analysis was conducted using ATLAS.ti (ATLAS.ti Scientific Software Development GmbH). The feasibility and usability of the newly developed momentary assessment tool were tested for 28 consecutive days in 5 patients with endometriosis-related pain symptoms.

Results: Momentary assessment items contained questions concerning endometriosis symptoms, general somatic symptoms, psychological symptoms, contextual information, and the use of food and medication. A morning questionnaire on sleep and sexuality was included. In a pilot study, the patients considered the tool easy to use but time consuming. The average compliance rate of momentary assessments was 37.8% (106/280), with the highest completion rate during the first week (39/70, 56%). Therefore, it is advisable to use the ESM for a maximum of 7 days.
Conclusions: A new digital tool for endometriosis symptom assessment was developed using the ESM, which may help overcome the limitations of current retrospective questionnaires. After validation and testing, future studies will be planned to evaluate the use of this tool in a clinical setting in order to propose a personalized treatment plan for women with endometriosis.

(JMIR Form Res 2021;5(12):e28782) doi:10.2196/28782

KEYWORDS
endometriosis; pelvic pain; positive affect; negative affect; patient-reported outcome measure; focus groups; experience sampling method; momentary symptom assessment; mobile phone

Introduction

Background

Endometriosis is defined as an estrogen-dependent condition involving the endometrium-like tissue outside the uterus [1]. It is estimated to be prevalent among approximately 10% in women of reproductive age and up to 50% in women with chronic pelvic pain (CPP) or fertility problems [2,3]. Dysmenorrhea, CPP, dyspareunia, fatigue, and infertility are the leading symptoms [4,5], which have a significant social and psychological impact, decreasing the quality of life of the patients [6-8]. Furthermore, the annual economic burden of women with endometriosis in European countries is high and similar to that of other chronic conditions [8]. The severity of the disease, as well as pelvic pain, infertility, and a higher number of years since diagnosis, are associated with higher costs of societal relevance given that these symptoms affect physical, mental, sexual, and social well-being, as well as work productivity [8-10].

Endometriosis is currently managed by surgical or medical interventions; however, approximately 50% of women have recurrent symptoms over a period of 5 years [11]. Moreover, the extent of endometriosis is not directly related to the degree of the symptoms [12], which suggests that the perception of symptoms may also be influenced by psychological and emotional distress [13-15].

Objective

Stratified and more individualized therapeutic approaches are needed to maximize treatment efficacy and improve physical, mental, sexual, and social well-being [8-10,16]. To do so, a reliable assessment of endometriosis-related symptoms is essential. Current guidelines for symptom assessment in patients with endometriosis include the recommendations of the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials [17] and the American Society for Reproductive Medicine [18]. The former has made recommendations for clinical outcomes in pain trials [17], including pain measured in 0 to 10 scales, physical functioning, emotional functioning, symptoms, and adverse events. For patients with endometriosis, the American Society for Reproductive Medicine [18] recommends daily ratings of pelvic pain, daily ratings of dysmenorrhea, and the Endometriosis Health Profile-30 (EHP-30) [19]. Currently, there is no available assessment tool for all contextual factors that could influence endometriosis complaints, including symptom triggers and overlapping symptoms with other comorbidities. Furthermore, validated questionnaires such as the EHP-30 are retrospective. The experience sampling method (ESM) is an electronic questioning method characterized by randomly repeated self-reports on symptoms, activities, emotions, or other elements of real-time daily life [20]. This momentary assessment method has several advantages, including the ability to assess the temporal relationship between variables, high ecological validity, and highly detailed information on the experiences of the subjects. This method aims to provide self-insight, personalized treatment approaches, and adequate monitoring of the effectiveness of these treatments in individual patients. Usually, this method is made available by the use of a mobile app [20-22].

Following the previous development of an ESM tool for psychiatric conditions [20], irritable bowel syndrome (IBS) [22-24], functional dyspepsia [25], and overactive bladder syndrome [26], we aimed to develop an ESM assessment tool for patients with endometriosis.

Methods

Overview

This study was conducted between August 2018 and September 2019 and consisted of 5 consecutive phases: initial item selection, focus group interviews to consider input from the patients, critical evaluation through expert meetings, development of the smartphone app, and a pilot study to evaluate feasibility and usability. This study was approved by the Institutional Review Board Ethics Committee of the Maastricht University Medical Centre (MUMC+), Maastricht, the Netherlands (Ref 2018-0674; 2019-1069), and the Máxima Medical Centre, Veldhoven, the Netherlands (Ref 18.122; L19.048).

Phase I: Question Selection

In agreement with the guidelines of the Food and Drug Administration on patient-reported outcome measure (PROM) development, item selection for the questionnaire started with an initial draft on the basis of the literature of validated outcome measures [27,28]. ESM-specific items concerning psychological, social, and environmental factors were derived from previous ESM validation studies [20-23]. Disease-specific items concerning the quality of life, affective symptoms, and disease-specific symptoms were derived from validated retrospective questionnaires (the Short Form-36, EHP-30, European Quality of Life-5 Dimensions, Generalized Anxiety Disorder-7, Patient Health Questionnaire-9, and Gastrointestinal Symptom Rating Scale-IBS). A list was created with all potentially relevant items from these questionnaires. The phrasing of the items was adjusted to conform to the momentary aspects of ESM assessments. The complete list of items was...
discussed with a multidisciplinary expert team consisting of gynecologists, endometriosis experts, urologists, a psychiatrist, a gastroenterologist, and a representative of the Dutch endometriosis patient organization. All the items were discussed for potential relevance. In addition, the experts were asked in an open discussion whether there were any relevant items missing according to their field of expertise.

**Phase II: Focus Groups**

**Focus Group Recruitment**

Premenopausal patients with endometriosis (diagnosed by physical examination and imaging techniques or laparoscopy) aged ≥18 years were recruited by gynecologists from the ward of the outpatient gynecology department at the MUMC+ or the Máxima Medical Center. Furthermore, patients were recruited through advertisements on the Dutch endometriosis foundation website. Pregnant women and patients with any organic explanation for CPP besides endometriosis were not eligible for participation. Furthermore, participants had to be able to speak and understand written Dutch, as the focus groups were conducted in this language. Written informed consent was obtained from all participants before the study.

**Focus Group Organization**

The focus groups were conducted according to the international PROM development guidelines [27] and the literature on focus group interviews [28]. For each focus group, 6 to 10 patients were invited, and 90-minute sessions were scheduled. The focus groups were conducted in 2 meeting phases according to the focus group guidelines [28], with the guidance of a moderator (EB) and at least one assistant moderator (AL, MP). In the first meeting phase, an open discussion, the participants were instructed to bring forward every item they considered essential for use in a real-time symptom assessment tool. In the second meeting phase, all items derived from the initial draft instrument were discussed in a structured manner. The patients could confirm or criticize the item value for momentary assessments and discussed the phrasing of the questions and the answer options. The focus groups were scheduled one meeting by one until saturation of input was reached, that is, the moment that the meetings no longer contributed any new items or information [25,26].

**Statistical Analysis**

The focus group discussions were voice-recorded and transcribed (JM). Data were qualitatively reviewed and systematically analyzed using ATLAS.ti software (ATLAS.ti Scientific Software Development GmbH; workbench for the qualitative analysis of large bodies of data, eg, textual, audio, and video). Each item was grouped by domain, and all domain items were clustered. When synonyms of items were used, the most frequently mentioned item was selected for the questionnaire.

**Phase III: Expert Meeting**

A final meeting with a multidisciplinary expert team (Phase I: Question Selection) was arranged to select the items to be used in the final questionnaire. The primary goal of the expert meeting was to critically discuss and convert the findings from the focus groups to generate applicable questions for clinical practice. A second goal was to shorten the list of ESM items to minimize response fatigue and, therefore, noncompliance of patients. All items that were included after the ATLAS.ti analysis of the focus group data were discussed for relevance until a majority was reached. In addition, the experts were asked in an open discussion whether there were any relevant items missing according to their field of expertise (ie, urology, gastroenterology, psychiatry, and gynecology).

**Phase IV: Development of a Smartphone App**

The smartphone app MEASuRE (Maastricht Electronic Abdominal Symptom Reporting) was previously created by MEMIC, the center for data and information management at the Faculty of Health, Medicine, and Life Sciences of the Maastricht University and the MUMC+. The app can measure real-time experiences in daily life using the concept of ESM. MEASuRE has been described in previous research and has been adjusted for patients with endometriosis using the questions that were selected in the final expert meeting [22-26].

**Phase V: Pilot Study**

The usability of the MEASuRE app has been thoroughly tested in patients with IBS. However, as we adapted the questions to an endometriosis-specific tool, we decided to conduct a pilot study with 5 patients with endometriosis to test the feasibility and usability of these changes to the tool. Given that endometriosis symptoms fluctuate during the menstrual cycle, we aimed to test whether collecting ESM data for 28 consecutive days was feasible [4,10]. Premenopausal women aged at least 18 years and diagnosed with endometriosis were recruited via the ward of the outpatient gynecology department of the MUMC+ or the Máxima Medical Center. The inclusion and exclusion criteria were similar to those in phase II, and written informed consent was obtained before participation. During the study period, ESM assessments were conducted on the patients’ smartphones using the MEASuRE app. Because the sampling procedure should cover a range of waking hours and activities, the momentary assessments started after 7:30 AM and finished before 10:30 PM. The app sent out a notification at 10 random moments during the day, each within a 90-minute time frame, after which the patients could complete the identical electronic self-reports. To minimize the extent to which data were influenced by retrospective biases, the participants had to respond to the notification within the requested time frame (10 minutes). After this period, which has also been described in other studies [20,29], it was no longer possible to start the assessment. Past research has typically used 5 to 10 assessments per day to measure real-time experiences in daily life [29,30]. As missing entries were expected, we also analyzed the rates of compliance of at least 3 out of 10 assessments each day. The participants were called on the second study day to check for technical difficulties and to ensure that the questions were clear. The patients were called and interviewed after 2 weeks and at the end of the pilot study to collect feedback concerning the logistics, usability, and content of the questionnaire.
Results

Phase I: Question Selection

Figure 1 systematically describes the development of a momentary PROM. During question selection, 54 items concerning psychological, social, and environmental factors were derived from questions used in previous ESM validation studies [20-23], whereas 30 items were derived from validated retrospective questionnaires (the Short Form-36, EHP-30, European Quality of Life-5 Dimensions, Generalized Anxiety Disorder-7, Patient Health Questionnaire-9, and Gastrointestinal Symptom Rating Scale-IBS) and made suitable for momentary assessment. Seven questions regarding physical and endometriosis-specific symptoms were added through a clinical literature search [1-5]. During the expert meeting, 13 items were excluded on the basis of relevance. Validated scales such as the Bristol Stool Chart (used in the ESM tool for patients with IBS) and a urological urgency scale were added to make it possible to compare data from patients with endometriosis and patients with other chronic abdominal pain [31,32]. A total of 78 ESM questions were selected concerning different domains: endometriosis-specific symptoms, general somatic symptoms, sleep, sexuality, mood and psychological factors, social and contextual factors, and use of nutrition and medication.

Figure 1. Process of patient-reported outcome measure development. ESM: experienced sampling method; PROM: patient reported outcome measure.

Phase II: Focus Groups

Overview

The characteristics of the women who participated in the focus group meetings are summarized in Table 1. A total of 19 patients initially agreed to participate in the focus groups; however, only 14 were present. The reasons for cancelation were illness (n=2), other plans (n=1), and family-related issues (n=1). One patient did not report any reason for the cancelation. The age of the participants ranged from 23 to 41 years. Saturation of input was reached after 3 focus groups. After the ATLAS.ti analysis of focus group data, the questionnaire comprised 56 items.
Table 1. Baseline characteristics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Study group</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Focus groups (n=14)</td>
<td>Pilot study (n=5)</td>
<td></td>
</tr>
<tr>
<td><strong>Sociodemographic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>37.1 (6.7)</td>
<td>35.6 (5.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Level of education, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>1 (7)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>College or university</td>
<td>13 (93)</td>
<td>5 (100)</td>
<td></td>
</tr>
<tr>
<td><strong>Occupational status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>1 (7)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>3 (21)</td>
<td>2 (40)</td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>10 (72)</td>
<td>3 (60)</td>
<td></td>
</tr>
<tr>
<td><strong>Relationship status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>1 (7)</td>
<td>1 (20)</td>
<td></td>
</tr>
<tr>
<td>In relationship</td>
<td>13 (93)</td>
<td>4 (80)</td>
<td></td>
</tr>
<tr>
<td><strong>Anthropometric</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m²), mean (SD)</td>
<td>27.6 (4.7)</td>
<td>26.5 (5.7)</td>
<td></td>
</tr>
<tr>
<td><strong>Medical, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of hormonal medication</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral contraceptives</td>
<td>12 (86)</td>
<td>4 (80)</td>
<td></td>
</tr>
<tr>
<td>Mirena IUD&lt;sup&gt;a&lt;/sup&gt;</td>
<td>4 (29)</td>
<td>2 (40)</td>
<td></td>
</tr>
<tr>
<td>Progestins</td>
<td>1 (7)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>GnRH&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4 (29)</td>
<td>2 (40)</td>
<td></td>
</tr>
<tr>
<td>Regular use of pain medication</td>
<td>11 (79)</td>
<td>5 (100)</td>
<td></td>
</tr>
<tr>
<td>Surgery for endometriosis</td>
<td>11 (79)</td>
<td>3 (60)</td>
<td></td>
</tr>
<tr>
<td>Infertility</td>
<td>5 (36)</td>
<td>1 (20)</td>
<td></td>
</tr>
<tr>
<td>Use of psychiatric medication</td>
<td>2 (14)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Traumatic life event in past</td>
<td>3 (21)</td>
<td>1 (20)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>IUD: intrauterine device.

<sup>b</sup>GnRH: gonadotrophin-releasing hormone.

**Morning Questionnaire**

Women with endometriosis and deep dyspareunia have been found to have lower sexual quality of life, presenting with impaired sexual functioning and decreased satisfaction, which, in turn, can negatively affect personal relationships [33]. Questions concerning sexual activity or avoidance were adapted from the modular dimension Sexual intercourse of the EHP-30. Furthermore, the patients considered questions regarding sleep relevant to the general state of well-being. These questions were added to the morning questionnaire, as it was considered unnecessary to assess these items repeatedly during the day [34].

**Momentary Assessments**

Most of the answer options were presented in the numeric rating scale from 0 to 10. However, some questions had answer options on a scale of −5 to +5. The list of questions was shortened by creating subquestions in the case of positive answers. In this matter, questions regarding sexual intercourse, urination, and defecation were asked retrospectively to check whether or not they occurred. If complaints arose, the follow-up questions were asked. The patients stated that the extent of vaginal blood loss was an important issue; however, they also noted that, in the case of absence of a menstrual cycle or after hysterectomy, they did not like to answer any questions regarding blood loss. This was solved by creating a one-off questionnaire on the menstrual cycle after downloading the app. The general somatic questions concerned symptoms as part of a psychosomatic syndrome or caused by the side effects of medication. Questions regarding psychological components were added. These questions concerned both negative and positive affect [20,35,36]. Social and contextual items were added because they could influence physical and emotional well-being and, therefore, the severity of the complaints [20,37]. Questions regarding food intake, use
of pain medication, and alcohol consumption were considered essential for influencing pain symptoms or general well-being.

**Phase III: Expert meeting**

During the final expert meeting, 6 items were excluded and 1 item was added. The question “How many times did you wake up last night?” was excluded on the basis of relevance, as the quality of sleep and the reason for waking up had already been assessed. In the psychological items, the questions *I feel lonely* and *I feel insecure* were excluded to shorten the list. Furthermore, the ATLAS.ti analysis revealed that these items were mentioned less frequently by patients. Three questions with synonyms regarding energy level (feeling tired, feeling lifeless, and feeling energetic) were adapted to 1 question. The final questionnaire consisted of 51 items (Figure 1). The domains defined during the question selection phase were retained. The number of ESM items varied depending on the answers given by the patients. A morning questionnaire comprised a minimum of 4 and a maximum of 7 questions and included information about sleep and sexuality. Momentary assessments regarding the remaining domains comprised a minimum of 31 and a maximum of 42 items. Table 2 shows the number of questions per category. Two questions were added to a one-off questionnaire on the menstrual cycles of the patients.

### Table 2. Number of experience sampling method (ESM) questions per category.

<table>
<thead>
<tr>
<th>Category</th>
<th>Maximum number of ESM questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-off questionnaire</td>
<td></td>
</tr>
<tr>
<td>Menstrual cycle</td>
<td>2</td>
</tr>
<tr>
<td>Morning questionnaire</td>
<td></td>
</tr>
<tr>
<td>Sleep</td>
<td>4</td>
</tr>
<tr>
<td>Sexuality</td>
<td>3</td>
</tr>
<tr>
<td>Momentary assessment</td>
<td></td>
</tr>
<tr>
<td>Endometriosis-specific symptoms</td>
<td>15</td>
</tr>
<tr>
<td>General somatic symptoms</td>
<td>7</td>
</tr>
<tr>
<td>Mood and psychological factors</td>
<td>7</td>
</tr>
<tr>
<td>Social and contextual factors</td>
<td>8</td>
</tr>
<tr>
<td>Use of nutrition and medication</td>
<td>5</td>
</tr>
</tbody>
</table>

**Phase IV: Development of a Smartphone App**

The final questionnaire that was built into the smartphone app MEASuRE is listed in English in Multimedia Appendix 1. This questionnaire was originally created in Dutch and was officially translated by Medilingua translations; however, it has not yet been validated in English.

**Phase V: Pilot Study**

**Feasibility and Compliance**

The characteristics of the women who participated in the pilot study are summarized in Table 1. The morning questionnaire took an average of 22 seconds to complete (range 11-44 seconds), and the momentary assessments took an average of 3 minutes and 2 seconds to complete (range of 72-255 seconds). The average completion rate for the morning questionnaires was 81% (23/28 study days). The average response rate for all momentary assessments was 37.86% (530/1400 questionnaires), with a range of 6.1% (17/280) to 56.1% (157/280) between patients. The average completion rate for a minimum of 3 questionnaires was 68% (19/28 study days). The response rate was highest during the first week of the pilot study, on average, 56% (39/70) of questionnaires, with a range of 21% (15/70) to 79% (55/70) between patients. The first week was the only week in which all participants completed at least 3 questionnaires on each study day. Figure 2 shows a histogram with the mean number of completed beep questionnaires (momentary assessments) per study day. In total, 0.79% (11/1400) of the momentary assessments were started but not completed.
**Figure 2.** Mean completed momentary assessments per study day. The horizontal dotted line represents the minimum of 3 completed questionnaires per day that is needed for a reliable analysis.

**Interview**

During interviews, the patients noted that the app was easy to use and that the questions were clearly defined, although some suggestions were made for clarifications or answer options. Table 3 shows the results of the interviews with patients and includes advantages and limitations. Recommendations from these patients were also included, and their recommendations concerning the content were added to the final questionnaire (Multimedia Appendix 1). There were no particular questions that the patients did not want to answer, such as questions concerning sexuality.

**Table 3.** Advantages, limitations, and recommendations mentioned by the patients that completed the pilot study (n=5).

<table>
<thead>
<tr>
<th>Theme</th>
<th>Advantages</th>
<th>Limitations</th>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usability</td>
<td>• “Easy to use.”</td>
<td>• “Loading of the questionnaire was slow in case of a bad internet connection.”</td>
<td>• Measure period for a maximum of 7 days:</td>
</tr>
<tr>
<td></td>
<td>• “Completing a questionnaire was getting easier and faster over time.”</td>
<td></td>
<td>• “During a menstrual period” (n=1)</td>
</tr>
<tr>
<td></td>
<td>• “During a menstrual period” (n=1)</td>
<td></td>
<td>• “Let the patient decide when to start measuring (when the complaints are highest)” (n=3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• “Measure period for one month to be sure that one whole menstrual cycle is included. Use less questionnaires per day to improve compliance.” (n=1)</td>
</tr>
<tr>
<td>Content</td>
<td>• “Very complete.”</td>
<td>• See recommendations for missing items according to the pilot study participants.</td>
<td>• “The option I have pain while laying down is missing.”</td>
</tr>
<tr>
<td></td>
<td>• “The questions are very clear.”</td>
<td></td>
<td>• “I am together with my pet.”</td>
</tr>
<tr>
<td></td>
<td>• “Only by filling in my symptoms and activities at the same time makes me aware of a symptom pattern.”</td>
<td></td>
<td>• “The activity: taking care of my children/family.”</td>
</tr>
<tr>
<td></td>
<td>• “This is the only questionnaire I know that also contains bowel and bladder symptoms.”</td>
<td></td>
<td>• “Add LEFT (L) and RIGHT (R) to the abdominal pain figure.”</td>
</tr>
<tr>
<td>Compliance</td>
<td>• “Timing ten times a day is good because then I don’t feel guilty when I miss a questionnaire.”</td>
<td>• “Work makes it difficult to complete the questionnaires.”</td>
<td>• “Completing 5-7 assessments per day is feasible.” (n=3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Social activities make it difficult to complete the questionnaire.”</td>
<td>• “Completing 4-5 assessments per day is feasible.” (n=2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Less motivation to fill in questionnaires when I don’t have any somatic complaints.”</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Less motivation to fill in questionnaires when I am feeling down.”</td>
<td></td>
</tr>
</tbody>
</table>
Discussion

Overview

Following the development of an ESM tool specific to psychiatric conditions [20] and gastrointestinal and urological disorders, such as IBS, functional dyspepsia, or overactive bladder syndrome [22-26], we developed a modern assessment tool for patients with endometriosis. This new tool was developed according to the international guidelines on PROM development and comprised 5 phases: a selection of items on the basis of a literature review, a focus group study, expert meetings, the development of an electronic PROM using a smartphone app, and testing of the usability and feasibility with a pilot study. During interviews, the patients noted that the app was easy to use and that the questions were clearly defined. During our pilot study, only 0.79% (11/1400) of all momentary assessments were started but not completed, indicating that the assessments were easy to complete and not too time-consuming. However, completing up to 10 momentary assessments each day was considered time-consuming and caused response fatigue and noncompliance. During a study period of 28 days, most assessments were completed during the first week (39/70, 56%, in the first week vs on average 106/280, 38%, during the total study period). Compared with other ESM studies, this compliance rate is relatively low, as meta-analyses have shown completion rates of 82% to 85% [38,39]. However, comparing data with other ESM studies is difficult because the absence of methodological guidelines related to the use of this method has resulted in a large heterogeneity of designs [39], and compliance rates have not been reported in approximately half of the studies [40]. For better compliance, fewer study days, less assessments per day, and fewer items per assessment are advised [38,40].

In addition, as previous ESM studies recommend at least 3 completed questionnaires per day for a reliable analysis, which occurred consistently only during the first week of this study, we recommend using the ESM for a maximum of 7 days [41]. However, as endometriosis can fluctuate during the menstrual cycle, assessing patients for 4 weeks could add valuable information and might be considered with fewer assessments per day.

Strengths and Limitations of the ESM

The ESM has several advantages over traditionally used assessment tools, including the ability to evaluate the temporal relationship between variables, high ecological validity, and highly detailed information on the experience of the subject. Furthermore, the ESM allows for a prospective, individualized within-person approach to symptoms and symptom formation and to treatment outcome, which contrasts with the average patient approach of traditional evidence-based practice [37,41]. Self-reports across multiple days and among various participants provide profound and comprehensive insights into the disease course and treatment efficacy. On the basis of this, the ESM may also provide clues for behavioral interventions, adding value to fragmented monodisciplinary treatment, which remains refractory to responsiveness.

A limitation of the ESM is that it is perceived as time-consuming and requires considerable motivation on the part of the patient. Therefore, assessments are ideally kept as brief as possible. Furthermore, assessments several days in a row could encourage rumination. Thus, on the basis of the recommendations of previous studies using ESM, we suggest limiting the assessment period to 7 days and adding items concerning positive affect [20,35,36]. Another concern is selection bias. Not all patients are willing to participate or comply with study protocols using ESM, and participation could be affected by motivation for change in treatment. However, previous research has shown that this method is feasible for a wide variety of patients [42,43].

Strengths and Limitations of PROM Development

Given that the questions in this new tool are derived from validated questionnaires, this ESM tool designed for use in patients with endometriosis is comparable to validated retrospective PROMs. The use of patient focus groups according to the international guidelines on PROM development strengthens the validity of the questionnaire. A limitation of our focus group study was the limited number of patients who participated. Although 19 patients agreed to participate, only 14 were included in the 3 focus groups. Ideally, 6 to 10 participants were scheduled for each focus group. Most importantly, saturation of input was reached. During the pilot study, a few recommendations were made regarding the content, and these were added to the final questionnaire (Multimedia Appendix 1).

Future Study Perspectives

This paper comprises the development (part I) of a new PROM for women with endometriosis, with the ability to assess symptoms in real time. The validation stage (part II) will involve testing the psychometric properties of this newly developed tool. A 7-day validation study will be conducted to assess content validity and to investigate the association with potential triggers of physical symptoms, such as psychological, social, and contextual factors. In the planned validation study, 25 patients with endometriosis with CPP at least 1 day per week on average will be included. By letting patients start measuring at random moments, we expect to collect enough data from different menstrual cycle phases and that there will be sufficient data after the use of ESM in 7 consecutive days. Data from this newly developed ESM tool will be compared with frequently used validated (retrospective) outcome measures such as the EHP-30 questionnaire and end-of-day and end-of-week retrospective pain scores. After validation and testing, future studies will be planned to evaluate the use of this tool in a clinical setting in order to propose a personalized treatment plan.

In conclusion, in agreement with the international guidelines, we developed a PROM for real-time symptom assessment in women with endometriosis. This new electronic tool consists of a morning questionnaire and momentary assessments with questions regarding physical, mental, sexual, and social well-being. This tool was considered easy to use and may help overcome the limitations of existing retrospective questionnaires. To minimize noncompliance, it is advised to use this tool for a maximum of 7 days.
Conflicts of Interest

GVK performs clinical trials with Medtronic, Boston Scientific, and Astellas and is a consultant to Medtronic, Boston Scientific, and Solace therapeutics. None of these conflicts are relevant to this work.

Multimedia Appendix 1

Set of questions for the endometriosis-specific experience sampling method–patient-reported outcome measure after focus groups, expert meetings, and pilot study.

References


Abbreviations
- CPP: chronic pelvic pain
- EHP-30: Endometriosis Health Profile-30
- ESM: experience sampling method
- IBS: irritable bowel syndrome
- MEASuRE: Maastricht Electronic Abdominal Symptom Reporting
- MUMC+: Maastricht University Medical Centre
- PROM: patient-reported outcome measure
Assessing the Views of Professionals, Patients, and Care Partners Concerning the Use of Computer Tools in Memory Clinics: International Survey Study

Aniek M van Gils1, MSc, MD; Leonie NC Visser1,2, PhD; Heleen MA Hendriksen1, MSc; Jean Georges3, AB; Majon Muller4, MD, Prof Dr; Femke H Bouwman5, MD, PhD; Wiesje M van der Flier1,5, Prof Dr; Hanneke FM Rhodius-Meester1,4, MD, PhD

1Department of Neurology, Alzheimer Center Amsterdam, Amsterdam Neuroscience, Amsterdam UMC, Location VUmc, Amsterdam, Netherlands
2Department of Neurobiology, Care Sciences and Society, Division of Clinical Geriatrics, Center for Alzheimer Research, Karolinska Institutet, Stockholm, Sweden
3Alzheimer Europe, Luxembourg, Luxembourg
4Department of Internal Medicine, Geriatric Medicine Section, Amsterdam Cardiovascular Sciences Institute, Amsterdam UMC, Location VUmc, Amsterdam, Netherlands
5Department of Epidemiology and Data Science, Amsterdam UMC, Location VUmc, Amsterdam, Netherlands

Corresponding Author:
Aniek M van Gils, MSc, MD
Department of Neurology, Alzheimer Center Amsterdam, Amsterdam Neuroscience
Amsterdam UMC, Location VUmc
De Boelelaan 1118
Amsterdam, 1081 HZ
Netherlands
Phone: 31 204440685
Email: a.vangils@amsterdamumc.nl

Abstract

Background: Computer tools based on artificial intelligence could aid clinicians in memory clinics in several ways, such as by supporting diagnostic decision-making, web-based cognitive testing, and the communication of diagnosis and prognosis.

Objective: This study aims to identify the preferences as well as the main barriers and facilitators related to using computer tools in memory clinics for all end users, that is, clinicians, patients, and care partners.

Methods: Between July and October 2020, we sent out invitations to a web-based survey to clinicians using the European Alzheimer’s Disease Centers network and the Dutch Memory Clinic network, and 109 clinicians participated (mean age 45 years, SD 10; 53/109, 48.6% female). A second survey was created for patients and care partners. They were invited via Alzheimer Europe, Alzheimer’s Society United Kingdom, Amsterdam Dementia Cohort, and Amsterdam Aging Cohort. A total of 50 patients with subjective cognitive decline, mild cognitive impairment, or dementia (mean age 73 years, SD 8; 17/34, 34% female) and 46 care partners (mean age 65 years, SD 12; 25/54, 54% female) participated in this survey.

Results: Most clinicians reported a willingness to use diagnostic (88/109, 80.7%) and prognostic (83/109, 76.1%) computer tools. User-friendliness (71/109, 65.1%); Likert scale mean 4.5, SD 0.7), and increasing diagnostic accuracy (76/109, 69.7%; mean 4.3, SD 0.7) were reported as the main factors stimulating the adoption of a tool. Tools should also save time and provide clear information on reliability and validity. Inadequate integration with electronic patient records (46/109, 42.2%; mean 3.8, SD 1.0) and fear of losing important clinical information (48/109, 44%; mean 3.7, SD 1.2) were most frequently indicated as barriers. Patients and care partners were equally positive about the use of computer tools by clinicians, both for diagnosis (69/96, 72%) and prognosis (73/96, 76%). In addition, most of them thought favorably regarding the possibility of using the tools themselves.

Conclusions: This study showed that computer tools in memory clinics are positively valued by most end users. For further development and implementation, it is essential to overcome the technical and practical barriers of a tool while paying utmost attention to its reliability and validity.

(JMIR Form Res 2021;5(12):e31053) doi:10.2196/31053
Introduction

Background

Dementia is a major health problem worldwide, with its prevalence expected to rise to 75 million patients by 2050 [1]. A timely and accurate diagnosis is essential for providing adequate care and appropriate treatment [2]. Diagnosing Alzheimer disease dementia or another type of dementia can be challenging, as clinical presentations overlap and multiple pathologies often co-occur [3,4]. Furthermore, increasing biomarker availability creates the possibility of diagnosing the early stages of Alzheimer disease before the onset of dementia and paves the way for individual dementia risk estimation [5,6].

With the availability of many different diagnostic tests, clinicians have the difficult task of combining and interpreting all the test results to come to an accurate diagnosis and prognosis [1-10] and communicating these results to patients and care partners [5,11,12]. Despite the increasing number of diagnostic tools, uncertainty in the diagnosis and prognosis of dementia remains, complicating the process of clearly explaining the test results [11]. Recent research has shown that most patients and care partners greatly value precise information on diagnosis and prognosis [12,14]. However, clinicians are often reluctant to address these topics during consultations, leaving these informational needs unmet [12,15].

Currently, artificial intelligence solutions are rapidly being developed and can aid clinicians in addressing these challenges in several ways. Artificial intelligence–based computer tools for dementia diagnosis and prognosis have demonstrated diagnostic accuracy equal to that of clinicians’ performance. These tools support individual risk estimation and increase clinicians’ confidence in diagnosis and prognosis [16-19].

Web-based cognitive test tools have shown promising results, enabling cost-effective testing [20-22]. From other medical fields, such as oncology, we know that computer-based tools can also support the communication process, for example, by engaging patients and their families more actively in the diagnostic decisions or by supporting clinicians in the clear communication of results [23,24]. To date, the actual implementation of such computer tools in memory clinic practice has been limited [25,26].

Barriers and Facilitators

Several barriers to the acceptance and implementation of tools have been identified in different health care areas. The main concerns regarding computer tools are related to the physician–patient relationship: the fear of interfering with this relationship when using a tool and affecting patient communication. Furthermore, clinicians fear the disturbance of clinical work and the loss of clinical autonomy when using a tool. In addition, a time-consuming tool, a tool that does not fit into the workflow, complexity of a tool, and computer literacy have been frequently mentioned as barriers. On the other hand, good training before the use of a tool, user-friendliness, relevance, transparency, and reliability are stimulating factors in the use of a computer tool in clinical practice [27-33].

It is not known if the same barriers and facilitators apply to computer tools in memory clinics. The nature of the patient population—older adults with cognitive decline—and the vast number of diagnostic tests involved in the diagnostic process might lead to a different set of relevant barriers and facilitators. In addition, patients’ and care partners’ opinions regarding the use of computer tools by their clinicians might be a potential barrier to or facilitator of clinicians’ use of a tool.

Objective

Therefore, this study aims to understand preferences and identify the main barriers to and facilitators of using computer tools in the dementia workup from the perspectives of clinicians, patients, and care partners, that is, the end users.

Methods

Design

We conducted 2 surveys, 1 for clinicians and 1 directed at patients and care partners, both in the fall of 2020. In addition, to aid the interpretation of the survey results, we conducted an interactive panel session with clinicians. The study was approved by the medical ethical committee of the Amsterdam University Medical Center, Vrije Universiteit Medical Center, Amsterdam. All participants provided digitally informed consent.

Survey for Clinicians

Participants

Between July and October 2020, we invited clinicians from memory clinics in Europe via the European Alzheimer’s Disease Consortium and the Dutch Memory Clinic network (Nederlands Geheugenpoli Netwerk [NGN]) through emails that contained a link to participate in the web-based survey. Furthermore, we invited all participants during the annual NGN conference to share their thoughts in our web-based interactive panel session. The survey was created in the web-based survey tool Survalyzer (Survalyzer AG) [34] and translated into Dutch and English. The survey was adaptive; that is, certain questions were only conditionally displayed based on responses to other items. Furthermore, participants could scroll through the survey to edit their answers. The survey comprised 3 parts. In the first part, we collected background information (eg, age, gender, profession, and specialization). In the second part, we used a funneled method to examine the current opinions on computer tools and identify the barriers and facilitators. First, we asked if the clinician would be willing to use computer tools in general, after which we asked them to explain their opinions in an open-ended question. Subsequently, we provided them with a list of barriers and facilitators compiled from barriers and facilitators known from the existing literature [27,28,32], and we asked them which factors would stimulate or discourage
them from using a tool. Clinicians could complement this list with their own perspectives. We then asked them to rate the importance of these factors using the Likert scale (1=very unimportant, 2=unimportant, 3=neutral, 4=important, and 5=very important). Finally, we asked the participants how likely they were to use diagnostic and prognostic tools. In the third part, we explored clinicians’ opinions on additional computer tools, that is, web-based cognitive testing, communication support, and communication skills training.

**Interactive Panel Session**

During the NGN annual conference (held on the web on November 10, 2020), 2 authors (HFMRM and LNCV) presented the preliminary results of the survey. To help interpret these results, they asked all conference participants several in-depth questions using Mentimeter [35]. These questions were related to the importance of several factors that stimulate their trust in a tool; factors that would convince them of the usability, reliability, and validity of a tool; and the primary outcome measures of a tool.

**Survey for Patients and Care Partners**

**Participants**

Between July and October 2020, we invited a mixed memory clinic population comprising patients with subjective memory complaints (subjective cognitive decline [SCD]), mild cognitive impairment, and dementia and care partners to participate in the web-based survey. To be included, patients had to be able to understand the questionnaire in Dutch or English. In this study, care partner refers either to an informal caregiver or a close relative or friend of the patient who provides either or both emotional and practical support. A general link to the survey was sent via a newsletter and social media to the members (patients and care partners) of Alzheimer Europe and directly to the members of the Alzheimer’s Society United Kingdom. Next, we invited both patients and care partners from the Amsterdam Dementia Cohort of the Alzheimer Center Amsterdam [36,37] and the Amsterdam Aging Cohort of the outpatient geriatric clinic of the Amsterdam University Medical Center (Amsterdam University Medical Centers) [38]. Patients and care partners were approached and informed by phone or email, and when they confirmed their participation, they were sent a personalized link to the web-based survey. The survey was adaptive to reduce the number of questions. Furthermore, participants could scroll through the survey to edit their answers.

**Survey**

A total of 2 versions of the survey were created, 1 directed at patients and 1 at care partners, both comprising 3 parts. In part 1, we collected background information regarding the participants (eg, age, gender, and diagnosis). In part 2, we asked for their opinion on clinicians’ use of (1) a computer tool that analyzes the results of the diagnostic tests (diagnostic tool), (2) a computer tool to help predict the course of their symptoms (prognostic tool), and (3) a tool to help communicate the test results in day-to-day language with the patient (communication tool). We adjusted the predefined list of barriers to and facilitators of using computer tools to the patient and care partner perspectives. We provided participants with this list and asked the extent (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, and 5=strongly agree) to which the different items applied to them. In the last part, we asked their opinion on computer tools directed at patients and care partners, that is, web-based cognitive testing and tools that could support and empower them in their communication with the clinicians. The survey was piloted in a test panel of 3 patients (2 with SCD and 1 with dementia) and 1 care partner of a person with dementia.

**Analysis**

Completion of the survey was enforced using mandatory questions. Proceeding with the survey was not possible when a question was unanswered. Furthermore, it was not possible to complete the survey more than once, as Survalyzer solely allowed unique visitors. Only completed surveys were analyzed. We analyzed participant characteristics and survey outcomes using descriptive statistics. Chi-square tests were used to compare answers between patients and care partners. For clinicians, we compared answers between groups based on age, sex, profession, and specialization. When using the 5-point Likert scale, the mean Likert scale scores were calculated per item. Frequencies were calculated for all the barriers and facilitators from the predefined list. We combined the frequencies with mean Likert scale scores to define the most important barriers and facilitators (eg, the item with the highest frequency combined with the highest mean Likert scale score was regarded as the most important). SPSS, version 22.0 (IBM Corporation) was used to analyze the quantitative data. P values <.05 were considered significant.

The answers to the open-ended questions were analyzed in MAXQDA software (VERBI Software) [39] using a process of deductive thematic content analysis [40,41]. A total of 2 authors, AMVG (physician) and HMAH (neuropsychologist), independently generated the initial thematic codes based on the existing literature and data. Subsequently, 1 author (AMVG) generated a thematic framework and used this framework to code all the given answers. The codes were then sorted into broad categories.

**Results**

**Demographics**

Sample descriptions have been presented in Table 1 for clinicians and Table 2 for patients and care partners. Clinicians were, on average, aged 45 (SD 11) years and had 16 (SD 13) years of experience. Most participating clinicians were medical specialists working in neurology (60/109, 55%) or internal or clinical geriatric medicine (33/109, 30.3%). Patients were in general older (mean age 73 years, SD 8) than care partners (mean age 65 years, SD 12), who were mostly a partner or spouse (33/46, 72%) or a grandchild, daughter, grandson, and son (in-law; 12/46, 26%). Participating patients were most often diagnosed with SCD (21/50, 42%), whereas participating care partners were mostly those of patients with dementia (36/46, 78%).
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Web-based survey (n=109)</th>
<th>Interactive panel session&lt;sup&gt;a&lt;/sup&gt; (n=184)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>45 (11)</td>
<td>43 (11)</td>
</tr>
<tr>
<td>Sex (female), n (%)</td>
<td>53 (48.6)</td>
<td>98 (85.9)</td>
</tr>
<tr>
<td><strong>Cohort, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>European Alzheimer’s Disease Consortium</td>
<td>53 (48.6)</td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Dutch Memory Clinic network</td>
<td>56 (51.4)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Profession, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD&lt;sup&gt;c&lt;/sup&gt;, specialist</td>
<td>87 (79.8)</td>
<td>60 (54.5)</td>
</tr>
<tr>
<td>MD, specialist training or not in specialist training</td>
<td>12 (10.9)</td>
<td>1 (0.9)</td>
</tr>
<tr>
<td>Physician assistant, nurse specialist, or specialized nurse</td>
<td>3 (2.8)</td>
<td>23 (20.9)</td>
</tr>
<tr>
<td>Neuropsychologist or psychologist</td>
<td>6 (5.5)</td>
<td>16 (114.5)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (0.9)</td>
<td>10 (9.1)</td>
</tr>
<tr>
<td>Experience&lt;sup&gt;d&lt;/sup&gt; (years), mean (SD)</td>
<td>16 (13)</td>
<td>N/A</td>
</tr>
<tr>
<td>**Specialization&lt;sup&gt;e&lt;/sup&gt;, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neurology</td>
<td>60 (50.4)</td>
<td>N/A</td>
</tr>
<tr>
<td>Clinical geriatric or internal medicine</td>
<td>33 (30.3)</td>
<td>N/A</td>
</tr>
<tr>
<td>Nursing home physician or general practitioner</td>
<td>2 (1.8)</td>
<td>N/A</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>9 (8.3)</td>
<td>N/A</td>
</tr>
<tr>
<td>Other</td>
<td>9 (8.3)</td>
<td>N/A</td>
</tr>
<tr>
<td>**Institution&lt;sup&gt;f&lt;/sup&gt;, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic or university hospital</td>
<td>68 (62.4)</td>
<td>N/A</td>
</tr>
<tr>
<td>Nonacademic teaching hospital</td>
<td>32 (29.4)</td>
<td>N/A</td>
</tr>
<tr>
<td>Nonteaching hospital</td>
<td>8 (7.3)</td>
<td>N/A</td>
</tr>
<tr>
<td>Mental health service</td>
<td>2 (1.8)</td>
<td>N/A</td>
</tr>
<tr>
<td>Other</td>
<td>3 (2.8)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

<sup>a</sup>Owing to the hybrid conference setting, not every description was available for all participants. For sex, n=114 participants replied to the question. For profession, n=110 participants replied to the question.<br><sup>b</sup>N/A: not applicable.<br><sup>c</sup>MD: medical doctor.<br><sup>d</sup>Only applicable for medical specialists.<br><sup>e</sup>Some clinicians had ≥1 specialization.<br><sup>f</sup>Some clinicians worked in ≥1 institution.
Table 2. Sample demographics of patients and care partners participating in the web-based survey (N=96).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Patients^a (n=50)</th>
<th>Care partners (n=46)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>73 (8)</td>
<td>65 (12)</td>
</tr>
<tr>
<td>Sex (female), n (%)</td>
<td>17 (34)</td>
<td>25 (54)</td>
</tr>
<tr>
<td><strong>Cohort, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alzheimer Europe or Alzheimer’s Society United Kingdom</td>
<td>2 (4)</td>
<td>14 (30)</td>
</tr>
<tr>
<td>Amsterdam dementia cohort</td>
<td>25 (50)</td>
<td>27 (52)</td>
</tr>
<tr>
<td>Amsterdam aging cohort</td>
<td>23 (46)</td>
<td>5 (18)</td>
</tr>
<tr>
<td><strong>Diagnosis^b, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCD</td>
<td>21 (42)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>MCI</td>
<td>16 (32)</td>
<td>8 (17)</td>
</tr>
<tr>
<td>Dementia</td>
<td>13 (26)</td>
<td>36 (78)</td>
</tr>
<tr>
<td><strong>Education^c, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1 (2)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Middle</td>
<td>22 (45)</td>
<td>16 (36)</td>
</tr>
<tr>
<td>High</td>
<td>26 (53)</td>
<td>27 (61)</td>
</tr>
</tbody>
</table>

^a Of these 50 patients, 20 (40%) completed the survey together with their care partner.
^b For the care partners, the numbers represent the diagnosis of their loved ones.
^c SCD: subjective cognitive decline.
^d MCI: mild cognitive impairment.
^e According to the Dutch Verhage scale (low 1-3; middle 5; high 6-7).

Survey of Clinicians

Opinions on the Use of Computer Tools

In response to whether they would be willing to use computer tools in their daily clinical practice, 51.4% (56/109) of clinicians said they would probably use a diagnostic tool, and 29.4% (32/109) said they would certainly use a diagnostic tool. Furthermore, 7.3% (8/109) said they would be unlikely to use a tool, 11.9% (13/109) answered neutrally, and none of the clinicians reported that they did not want to use a tool. The results were similar for prognostic tools; of the 109 clinicians, 53 (48.6%) said they would probably use a prognostic tool, and 30 (27.6%) said they would certainly use a prognostic tool. Furthermore, 2.8% (3/109) were unlikely to use a prognostic tool, and 0.9% (1/109) would certainly not use a prognostic tool. The remaining participants responded as neutral (22/109, 20.2%). We found no differences in both diagnostic and prognostic tools based on age (P=.20 and P=.49, respectively), sex (P=.14 and P=.73, respectively), or profession (P=.61 and P=.98, respectively). We found that neurologists indicated more willingness to use prognostic tools (P=.04) than clinicians from other specializations.

Content analysis of clinicians’ explanations of their opinion on the use of computer tools resulted in 6 main topics: support, clinical expertise, efficiency, accuracy, clinician–patient relationship, and care of the future. Each of the topics has been described in Table 3, and illustrative quotes have been provided.
### Table 3. Clinicians’ opinions on the use of computer tools in memory clinics.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Description</th>
<th>Facilitating factors</th>
<th>Hindering factors</th>
<th>Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support</td>
<td>Support to the diagnostic process from screening or prescreening to follow up, support data storage, support research purposes</td>
<td>Not applicable for specific patient populations</td>
<td>“[…] I would welcome a tool that would be implemented with the available clinical data and help reach a diagnosis (ie, considering the neuropsychological and neuromaging data, in patients with such profile, expert diagnosis would be…with a probability of…%—which could be increased by the use of…biomarker).” [Male, 42 years, MD(^b), physician working in neurology]</td>
<td></td>
</tr>
<tr>
<td>Clinical expertise</td>
<td>Complementary to clinical expertise (eg, an aid for complex cases or with interpretation of test results) and contributory to evidence-based medicine [42]</td>
<td>Tools should not be a replacement for clinical expertise</td>
<td>“Computer tools and AI might be a way to have an evidence-based standard procedure in addition to my own long time clinical experience.” [Female, 59 years, MD, geriatrician]; “[…] I consider the clinical view as most important. A computer tool cannot (partly) replace this.” [Female, 38 years, MD, geriatrician]</td>
<td></td>
</tr>
<tr>
<td>Efficiency</td>
<td>The ability to standardize the diagnostic process, if easy to use, if connecting with electronic patient file, and time-efficiency</td>
<td>A tool not connected to the electronic patient file, information technology issues</td>
<td>“A quick and useful way to get practical answers on the workplace.” [Male, 62 years, MD, neurologist]; “For tools that are not implemented in the electronic patient file I foresee barriers in the implementation.” [Female, 45 years, MD, geriatrician]</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>Computer tools could help in making a more accurate diagnosis, providing additional objective information, and overcoming human errors</td>
<td>Tools might generate results of no use and fear of loss of important clinical information</td>
<td>“Sometimes we can be influenced by the patient we have in front of us. We can diagnose them too easily or consider them as (sub)normal because their general behavior makes us think so. A computer could be more objective than we are in some cases.” [Male, 26 years, MD, neurology resident]; “[…] I am afraid that there will be an outcome that is of no use for me, such as 64% chance of Alzheimer’s disease.” [Female, 38 years, MD, geriatrician]</td>
<td></td>
</tr>
<tr>
<td>Clinician–patient relationship</td>
<td>Improving patient communication</td>
<td>A tool might have a negative impact on the relationship between clinicians and patients</td>
<td>“[…] It facilitates the communication to the patient.” [Female, 49 years, MD, neurologist]; “Patients also come for attention and care, which they get less if we look at the screen more often.” [Male, 32 years, MD, physician working in neurology]</td>
<td></td>
</tr>
<tr>
<td>Care of the future</td>
<td>The use of tools is considered part of the care of the future</td>
<td>N/A(^a)</td>
<td>“AI and big data are the future, they make the invisible visible […]” [Male, 33 years, MD, internal (geriatric) medicine resident]</td>
<td></td>
</tr>
</tbody>
</table>

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

\(^b\)MD: medical doctor.

### Barriers and Facilitators

Clinicians’ selection of factors that would hinder the use of computer tools has been displayed in Figure 1A. None of the barriers were rated using a mean Likert scale score ≥4. Figure 1B shows which factors would stimulate the use of computer tools; items marked with an asterisk were rated with a mean Likert scale score of ≥4. The most important barriers were (1) a tool not being well connected with the electronic patient file (47/109, 43.2% of clinicians), with a mean Likert scale score of 3.8 (SD 1.0), and (2) important information, such as observations, cannot be added to a tool (49/109, 44.9% of clinicians), with a mean Likert scale score of 3.7 (SD 1.2). The most important facilitators were (1) user-friendliness (73/109, 66.9% of clinicians; mean 4.5, SD 0.7), and (2) increasing diagnostic accuracy (78/109, 71.6% of clinicians; mean 4.3, SD 0.7).
Additional Tools for Memory Clinics

Many clinicians (48/109, 44%) reported a willingness to use both web-based communication tools and skills training. Over half of the clinicians (62/109, 56.9%) indicated that they would like to test the patient’s cognition via the computer at home before the appointment. Frequently selected reasons for web-based cognitive testing were triage (62/109, 56.9%) and shortening of the test battery at the clinic (37/109, 33.4%). Not being able to observe the patient during testing was mentioned by 89% (42/47) of the clinicians who would not want to test the patient’s cognition on the web.

Interactive Panel Session

Both transparencies about the objectives of the tool provider (Likert scale score mean 4.3, SD 1.1) and honesty about the possibilities and limitations of a tool (Likert scale scores mean 4.4, SD 0.9) were considered important to strengthen clinicians’ trust in a tool. The most important aspects of convincing clinicians of usability, reliability, and validity of a tool were the explicit provision of information regarding the tool (mean 4.3, SD 0.9), scientific publications on underlying models (mean 4.3, SD 0.9), obtaining hands-on experience with the tool (mean 4.2, SD 0.6), and a randomized controlled trial (RCT) to test the effectiveness of the tool in clinical practice (mean 4.0, SD 0.8). Diagnostic accuracy (31/100, 31%) and patient-related
outcome measures (38/100, 38%), such as quality of life, were most frequently selected as the ideal primary outcome measures of such an RCT.

**Survey of Patients and Care Partners**

**Tools Used by Clinicians**

The results of patients’ and care partners’ opinions on diagnostic, prognostic, and communication tools have been presented in Table 4. Most patients and care partners were positive regarding their clinician using these tools. No differences were found between patients and care partners (diagnostic tools, $P=.36$; prognostic tools, $P=.63$) or different syndrome diagnoses (diagnostic tools, $P=.64$; prognostic tools, $P=.69$; communication tools, $P=.92$).

**Table 4.** Opinion of patients and care partners on clinicians’ use of diagnostic, prognostic, and communication tools, illustrated with quotes (N=96).

<table>
<thead>
<tr>
<th>Opinion</th>
<th>Patients (n=50), n (%)</th>
<th>Care partners (n=46), n (%)</th>
<th>Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagnostic tool</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think that is a good thing</td>
<td>38 (76)</td>
<td>31 (67)</td>
<td>“The more information, the better. As long as the computer program is in addition to the doctor’s expertise and not a replacement, I think it would be a good idea.” [Female 60 years, care partner]</td>
</tr>
<tr>
<td>I would not want that</td>
<td>4 (8)</td>
<td>4 (9)</td>
<td>“I think face to face contact between the doctor and the patient is essential.” [Female 74 years, care partner]</td>
</tr>
<tr>
<td>I do not know or no opinion</td>
<td>8 (16)</td>
<td>11 (24)</td>
<td>“Depends on how good the program is.” [Male 76 years, patient, dementia]</td>
</tr>
<tr>
<td><strong>Prognostic tool</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think that is a good thing</td>
<td>41 (82)</td>
<td>32 (70)</td>
<td>“There is nothing against the use of a computer in predicting the disease process. It remains an aid to the physician. [...] He/she should remain leading.” [Male 78 years, patient, SCD]</td>
</tr>
<tr>
<td>I would not want that</td>
<td>4 (8)</td>
<td>6 (13)</td>
<td>“I want to know so I can plan ahead. However, with the variation in the progression rate, I don’t see how this could be sufficiently accurate. If not accurate, I would not want it.” [Female, 61 years, patient, dementia]</td>
</tr>
<tr>
<td>I do not know or no opinion</td>
<td>5 (10)</td>
<td>8 (17)</td>
<td>“My husband approves [the use of a prognostic tool], me as his wife, do not know if I would like it. What if the prediction is somber! We would instantly be depressed.” [Female (age unknown), care partner]</td>
</tr>
<tr>
<td><strong>Communication tool</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| I think that is a good thing | 39 (78) | 38 (83) | —
| I would not want that | 3 (6) | 1 (2) | —
| I do not know or no opinion | 8 (16) | 7 (15) | —

aSCD: subjective cognitive decline.
bNo quotes available.
Figure 2. Agreement of patients and care partners on several statements regarding computer tools on a 5-point Likert scale. * Items rated with a mean Likert scale score of ≥4.

**Tools Directed at Patients and Care Partners**

Most patients (35/50, 70%) and care partners (26/46, 57%) expressed a preference for a list of example questions to select questions they wanted to ask their clinician. Furthermore, 42% (21/50) of patients and 50% (23/46) of care partners said they would prepare their visit to the memory clinic by watching informational videos. A smaller but considerable proportion of care partners (11/46, 24%) indicated that they would use a web-based communication tool to practice their communication skills, and some (5/50, 10%) patients reported wanting to do the same. A small proportion of patients (4/50, 8%) and care partners (8/46, 17%) did not want to use any web-based communication tools.

Most patients (35/50, 70%) and care partners (28/46, 61%) were positive regarding web-based cognitive testing at home. Reasons for not wanting to perform cognitive testing at home differed between patients and care partners (P=.01). The most frequently selected reason for patients was their preference for personal contact with the clinician (18/26; 69%). The most selected reason (5/12, 42%) for care partners was that web-based cognitive testing was too difficult to perform for their loved ones.

**Discussion**

**Principal Findings**

This study showed that most clinicians, patients, and care partners were supportive of the use of computer tools in memory clinics. This holds true for diagnostic and prognostic tools, tools that support communication, and web-based cognitive testing. Despite acknowledging their potential barriers, the general attitude of clinicians toward these tools was positive. The facilitating factors were mainly practical (tools should be user-friendly) and technical (connection with electronic patient files) and that tools should increase diagnostic accuracy. The identified barriers mainly focused on doubts regarding reliability and validity, preservation of clinical autonomy, and fear of losing important clinical information. Furthermore, the tools should be considered in addition to the current working methods and not as a replacement.

We hypothesized that the barriers to and facilitators of tools in memory clinics might differ from those identified in other health care areas because of the nature of the patient population (older adults with cognitive decline) and the large number of diagnostic tests used in the clinical workup of dementia. However, the barriers and facilitators we found in this survey study largely corresponded to the existing literature on barriers to and facilitators of the use of computer tools in other medical fields [27-33]. Furthermore, we found that most of the given answers to the open-ended question were in accordance with our predefined list, which was based on studies on applying computer tools in other health care areas [27,28,32]. In addition, it is conceivable that patients’ and their care partners’ (negative) opinions regarding the tools were a possible barrier to clinicians using a tool. In this study, we showed that patients, despite their age and (potential) cognitive decline, are mainly positive regarding the use of computer tools. Most of them embrace the possibility of using a tool themselves, and their care partners share this opinion. A computer tool must support rather than replace clinicians, who would then view it as an aid appropriate to modern times.

**Increase Acceptance of Tools**

We found the overall attitude among clinicians toward tools to be highly positive. Nevertheless, none of the available tools are regularly used in daily practice, and it seems there is a major information gap and educational need to make clinicians understand the possibilities of such computer tools [43]. The results from our survey provide direction for the way to increase
the acceptance of computer tools in memory clinic practice. First, clinicians indicated that an RCT on the efficiency in clinical practice would boost their confidence in the reliability of computer tools. RCTs are considered important and robust methods for assessing the impact of a tool [44,45]. However, no RCT on the application of computer tools has yet been performed in memory clinics. Thus, our findings suggest a need for RCTs with diagnostic accuracy and patient-related outcome measures as primary outcome measures.

Second, computer tools could contribute to evidence-based medicine (EBM) [42]. EBM concerns medical practice based on the best available evidence, clinical experience, and patient preferences [42]. Within the concept of EBM, clinical experience is highly essential and should not be replaced by a tool. Computer tools could support EBM by making the best available evidence more accessible to clinicians, and computer tools could clarify patient preferences. Acknowledging computer tools as a part of EBM might lead to clinicians viewing these tools as an aid complementary to their own clinical experience rather than a threat to their clinical autonomy [27,29,32,46]. In addition, there must be scientific publications regarding the underlying models and the transparent provision of information regarding a tool’s reliability to increase clinicians’ trust in the tools [28,32,47].

Then, clinicians’ confidence in the tools might be strengthened if the tools are under the jurisdiction of a regulatory body to authorize and supervise the quality. To date, there are no formal regulatory standards for tools used to support clinicians in decision-making [28]. When developers claim that their software has a medical purpose, it becomes a medical device, and manufacturers themselves have to proclaim that their device meets the safety and performance requirements. The new European Medical Devices Regulation implemented in May 2021 makes manufacturers adhere to more strict guidelines for ensuring the safety of their products, including assessment of their device by a notified body. Involving an independent notified body in the approval of a tool might be the first step toward improving its transparency and acceptance. Eventually, as a next step, the use of software should be included in the guidelines of professional associations [26].

The attitude toward tools is one of the key characteristics of eventual acceptance [30,32]. Several user acceptance models have been proposed to further encourage the acceptance of tools in medical practices [32]. In this context, we would like to address the user acceptance and system adaptation design model [32]. This model aims to include end users as the central point in the design process of a computer tool. In this model, user expectations and needs need to be thoroughly understood before developing a tool. The development of a tool is an iterative process, and end users should be continuously involved throughout the development process. In congruence with this user acceptance model, we took the first step toward accepting computer tools by identifying the barriers to and facilitators of computer tools according to the end users in memory clinics. On the basis of the user acceptance and system adaptation design model, the next step in the iterative process would be the evaluation of usability. Then, a pilot study would be needed to evaluate user acceptance. The results of each step should direct improvements in the tool and be evaluated in the next step. However, it should be mentioned that in this study, we asked participants about the barriers to and facilitators of computer tools in general. Therefore, the barriers and facilitators found in this study should be considered a starting point. When implementing specific tools, further exploration of user expectations and needs for those specific tools might be necessary.

**Strengths and Limitations**

One of the strengths of this study is that we involved all the end users of computer tools in memory clinics, that is, clinicians, patients, and care partners. Including patients and care partners originating from Europe and a heterogeneous population of clinicians contributed to the generalizability of the results. Furthermore, we used a funneled method in which we started the survey with open-ended questions and worked toward closed questions. By doing so, we actively asked and stimulated clinicians’ own input.

Nevertheless, our study has several limitations. First, we distributed the survey via a web-based link, which might have caused selection bias by only involving people with sufficient digital skills who might have had a more positive approach toward computer tools. We tried to minimize this risk for patients and care partners by including participants from both geriatric and neurology departments with different cognitive impairment stages. Furthermore, participants’ ages ranged from relatively young to older patients, who might have had less experience with digital tools. Second, we have no data available on the origin of the European participants, which might have led to an uneven distribution of participants among countries. Nonetheless, the international character of this study contributes to the generalizability of the results. Third, there might be a risk of response bias; people who are less inclined to use computer tools may not have responded. We could not estimate the response rates as the link to the survey was spread among an unknown number of people. Nonetheless, we gained insight into the important barriers and facilitators based on inquiry among large samples of the most important stakeholders.

**Conclusions**

In conclusion, this study shows broad support for the use of computer tools in memory clinic practices by clinicians, patients, and care partners. To stimulate the implementation of tools in daily memory practice, the tools should overcome several technical and practical barriers. Moreover, clinicians have to be convinced regarding the reliability and validity of the tool. By identifying the potential barriers and facilitators, we have paved the way for further development and implementation of the tools. Our results provide an important step in the iterative process of developing computer tools for memory clinics in cocreation with end users.

https://formative.jmir.org/2021/12/e31053
Acknowledgments

Research of the Alzheimer’s Center’s Amsterdam branch is part of the neurodegeneration research program of Amsterdam Neuroscience. The Vrije Universiteit Medical Center Amsterdam Center is supported by the Stichting Alzheimer Nederland and Stichting Vrije Universiteit Medical Center Fonds. The clinical database structure was developed with funding from the Stichting Dioraphe. The collaboration project Dementia diagnoses using Artificial Intelligence (light weighted computer tool for local memory clinics; project number LSHM19123-HSGF) is cofunded by the public–private partnership allowance made available by Health-Holland, Top Sector Life Sciences and Health, to stimulate public–private partnerships. The chair of WMVDF is supported by the Pasman Stichting. WMVDF and LNCV are recipients of ABOARD (A personalized medicine approach for Alzheimer disease), which is a public–private partnership receiving funding from ZonMW (number 73305095007) and Health-Holland, Top Sector Life Sciences and Health (public–private partnership allowance; number LSHM20106). LNCV is supported by a fellowship grant received from Alzheimer Nederland (WE.15-2019-05). WF, LNV and JG are recipients of the EU Joint Programme - Neurodegenerative Disease Research (JPND) project EURO-FINGERs (ZonMW-Memorabel #733051102) which is supported through the following funding organizations under the aegis of Joint Programme-Neurodegenerative Disease: Finland, Academy of Finland; Germany, Federal Ministry of Education and Research; Spain, National Institute of Health Carlos III; Luxemburg, National Research Fund; Hungary, National Research, Development and Innovation Office; The Netherlands, Netherlands Organisation for Health Research and Development; Sweden, Swedish Research Council (grant agreement INTER/JPND/19/BM/14012609). These funding sources were not involved in the study design, collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Conflicts of Interest

HFMMR performs contract research for Combinotics; all funding is paid to her institution. WMVDF performs contract research for Biogen. Research programs of WMVDF have been funded by ZorgOnderzoek Nederland, area of Medical Sciences, Dutch Research Council, European Union Framework Programmes, Alzheimer Nederland, Cardiovascular Onderzoek Nederland, Gieskes-Strijbis fonds, Pasman Stichting, Boehringer Ingelheim, Piramal Neuroimaging, Combinotics, Roche BV, AVID. She has been an invited speaker at Boehringer Ingelheim and Biogen. All funding is paid to her institution. FHB performs contract research for Optina Dx and Optos, she has been an invited speaker at Roche and has been invited for expert testimony at Biogen. All funding is paid to her institution. All other coauthors report no conflicts of interest.

References


34. Survalyzer - to survey and analyze. Utrecht. URL: https://education.survalyzer.com/ [accessed 2020-09-11]


43. van der Flier WM, van Gils AM, Visser LNC, Hendriksen HMA, Georges J, Muller M. Subjective cognitive decline in the operating room. JMIR Hum Factors 2019 Apr 23;6(2):e12155 [FREE Full text] [doi: 10.2196/12155] [Medline: 32047862]

44. van der Flier WM, van Gils AM, Visser LNC, Hendriksen HMA, Georges J, Muller M. Subjective cognitive decline in the operating room. JMIR Hum Factors 2019 Apr 23;6(2):e12155 [FREE Full text] [doi: 10.2196/12155] [Medline: 32047862]


A Multimodal Messaging App (MAAN) for Adults With Autism Spectrum Disorder: Mixed Methods Evaluation Study

Mohamad Hassan Fadi Hijab¹, MSc; Dena Al-Thani¹, PhD; Bilikis Banire¹, PhD
Division of Information and Computer Technology, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

Corresponding Author:
Mohamad Hassan Fadi Hijab, MSc
Division of Information and Computer Technology
College of Science and Engineering
Hamad Bin Khalifa University
Qatar Foundation
Education City
Doha, 34110
Qatar
Phone: 974 50677321
Email: mhhijab@hbku.edu.qa

Abstract

Background: Individuals with autism spectrum disorder (ASD) often exhibit difficulties in social and communication skills. For more than 30 years, specialists, parents, and caregivers have used techniques, such as applied behavioral analysis, augmentative and alternative communication, and the picture exchange communication system to support the social and communication skills of people with ASD. Even though there are many techniques devised to enhance communication, these techniques are not considered in existing social media apps for people with ASD. Even though there are many techniques devised to enhance communication, these techniques are not considered in existing social media apps for people with ASD.

Objective: This study aimed to investigate the effect of adding accessibility features, such as text-to-speech (TTS), speech-to-text (STT), and communication symbols (CS), to a messaging app (MAAN). We hypothesized that these accessibility features can enhance the social and communication skills of adults with ASD. We also hypothesized that usage of this app can reduce social loneliness in adults with ASD.

Methods: Semistructured interviews were conducted with 5 experts working in fields related to ASD to help design the app. Seven adults with ASD participated in the study for a period of 10 to 16 weeks. Data logs of participants’ interactions with the app were collected. Additionally, 6 participants’ parents and 1 caregiver were asked to complete a short version of the Social and Emotional Loneliness Scale for Adults (SELSA-S) questionnaire to compare pre-post study results. The Mobile Application Rating Scale: user version questionnaire was also used to evaluate the app’s usability. Following the study, interviews were conducted with participants to discuss their experiences with the app.

Results: The SELSA-S questionnaire results showed no change in the family subscale; however, the social loneliness subscale showed a difference between prestudy and poststudy. The Wilcoxon signed-rank test indicated that poststudy SELSA-S results were statistically significantly higher than prestudy results ($z = -2.047; P = .04$). Point-biserial correlation indicated that the SELSA-S rate of change was strongly related to usage of the TTS feature ($r = 0.708; P = .04$) and CS feature ($r = -0.917; P = .002$), and moderately related to usage of the STT feature ($r = 0.428; P = .17$). Lastly, we adopted grounded theory to analyze the interview data, and the following 5 categories emerged: app support, feature relevance, user interface design, overall feedback, and recommendations.

Conclusions: This study discusses the potential for improving the communication skills of adults with ASD through special features in mobile messaging apps. The developed app aims to support the inclusion and independent life of adults with ASD. The study results showed the importance of using TTS, STT, and CS features to enhance social and communication skills, as well as reduce social loneliness in adults with ASD.

(JMIR Form Res 2021;5(12):e33123) doi:10.2196/33123

KEYWORDS

autism; assistive technology; mobile app; social and communication skills
**Introduction**

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that is characterized by repetitive patterns of behavior, restricted interest, and deficits in social and communication skills [1]. These deficits affect the mental development of people with ASD, especially their socioemotional development, which may result in social and family loneliness [2]. Loneliness can be described as a lack of a close or intimate attachment to family members and friends. The consequences of loneliness include feelings of emptiness, anxiety, depression, suicide, and anger, which are detrimental to both physical and mental health [3]. Another effect of loneliness is learning disabilities, as the deficits associated with ASD limit patients’ opportunities to interact with their social environment and result in delays in learning social norms [3]. Understanding social norms is essential for positive and long-lasting social relationships. For example, an individual with ASD might not know when someone is angry and how to reciprocate. This lack of social norms may lead to social rejection. Social and communication skills are thus crucial components for social and emotional development. In recent years, the ASD prevalence rate has increased, with significant variability worldwide. In the United States, 185 in 10,000 children have a diagnosis of ASD [4]. Similarly, recent prevalence studies in the Middle East reported that 114 in 10,000 children from Qatar [5], and between 49 and 513 in 10,000 children from Lebanon have the disease [6]. Given the increasing prevalence of ASD, the impact of problems in social and communication skills associated with this disorder may lead to a high societal cost of supporting the physical and mental development of individuals with ASD. Thus, it is critical to develop innovative communication platforms to enhance social and communication skills between individuals with ASD and others.

Many interventions have been used to enhance social and communication skills, such as applied behavioral analysis (ABA) [7], augmentative and alternative communication (AAC) [8], and the picture exchange communication system (PECS) [9]. ABA is known as the primary treatment for children with ASD [10]. It is considered necessary for education and behavioral interventions, as it is commonly used to assist and improve social and communication skills for children with ASD. AAC provides individuals with ASD with pictures and illustrations to express their needs, and it is extensively used, although some caregivers and specialists have expressed concerns regarding its application. They often highlight that usage of AAC will prohibit the development of verbal speech in individuals with ASD [11]. Nevertheless, these technologies have the potential for supporting the social and communication skills of individuals with ASD [12]. The number of studies using AAC has increased due to the rise in its demand and value for an assistive technology tool in the last few years [13]. Lastly, the PECS is considered as one of the forms of AAC to help individuals with ASD to communicate. These systems appear promising due to their ability to construct sentences using images and symbols [14], with a low requirement of minor motor planning skills and little cognitive demand, which is why limited training is needed [15]. Many studies have investigated the importance of the PECS and determined its major role in enhancing social and communication skills [14,16]. On the other hand, PECS has several disadvantages, including difficulties in virtual communication and low speed in constructing messages [15]. However, behavioral and educational interventions, and digital revolutions have tremendously influenced the lives of individuals with ASD positively [17]. These technologies have increased the utilization of virtual communication platforms and highlighted the importance of providing remote services to the population with ASD [18]. Moreover, technology can ease communication and enhance individuals’ social well-being [19].

This study intends to verify the effectiveness of a new mobile messaging app called “MAAN,” which means “together” in Arabic, for adults with ASD, using pre-post study data. The data were collected using questionnaires, interviews, and interaction logs. The aim of MAAN is to provide an accessible, safe, and easy-to-use messaging app for adults with ASD and hence support their inclusion in society. In this study, we hypothesized the following: (1) The accessibility features of mobile apps (text-to-speech [TTS], speech-to-text [STT], and communication symbols [CS]), which are applied to MAAN, can enhance the social and communication skills of adults with ASD; and (2) The usage of the app can reduce social loneliness in adults with ASD.

**Methods**

**Study Design**

A mixed methods approach was followed in this study. Seven adults with ASD were asked to use MAAN with their caregivers for 16 weeks. Prior to the design and development of MAAN, semistructured interviews were conducted with experts to discuss the preliminary proposed app design. The app design was then modified according to the received feedback. After that, adults with ASD and their parents or caregivers were asked to use MAAN for a period of 16 weeks. This duration was adopted from the study by Laugeson et al [21], which suggested that when receiving social and communication skills interventions, improvement is only evident at 16 weeks. The study started by contacting individuals willing to participate.Due to COVID-19, the initial meetings with the participants were conducted online. All participants received an email for explaining the study information, such as date and time, and a weblink for collecting informed consent for a virtual meeting with the participants, their friends, and their parents. After receiving the participants’ consent, they were asked to fill an online prestudy questionnaire in order to collect demographic information. The parents or caregivers of the participants were also asked to complete a short version of the Social and Emotional Loneliness Scale for Adults (SELSA-S) questionnaire [21]. The aim of this questionnaire was to determine the social and family loneliness of the participants. In a virtual meeting,
the principal researcher explained the app features in detail and answered further questions regarding the study. The participants were also provided with a hotline to report any issue they might face during this study. To ensure the continuity of data collection, the participants were asked to make sure that the app was constantly running in the background.

At the end of the study, all participants with ASD and their caregivers or parents completed the Mobile Application Rating Scale: user version (uMARS) [22] questionnaire. The uMARS is a simple and reliable tool that can be used by end-users to assess the quality of mobile health apps. It provides a 20-item measure that includes 4 objective quality subscales (engagement, functionality, esthetics, and information quality) and 1 subjective quality subscale. The purpose of this questionnaire is to measure the app’s quality in terms of usability. The participants’ parents or caregivers also completed the SELSA-S questionnaire again to determine the participants’ social and family loneliness rates after the study. Lastly, a semistructured interview with parents or caregivers was conducted to gather information about the context of app usage. The interview included questions about the overall app experience in terms of usability, functionality, and challenges encountered while using the app, as well as suggestions for further enhancement.

MAAN App Design and Development

The app consisted of the following 3 main features: TTS, STT, and CS. One of the well-known features that can also be considered as a type of assistive technology is the TTS feature, which allows users to listen to written text on a computer, tablet, or smartphone. This technology is popular among people having difficulties in reading and decoding words [23]. Similarly, STT is an assistive technology feature used by people who struggle with writing. This form of assistive technology allows speech to be converted into text using the computing device. Both features are very commonly employed to support inclusion education settings [24]. The app also included the CS feature, which is an essential part of AAC-supported technologies, and its primary purpose is to assist individuals with ASD in constructing sentences. We are not aware of any messaging app that supports the use of the CS feature. “TAWASOL symbols” were employed for the CS feature. They were developed in 2013 by Mada Center, an assistive technology center in Qatar. These symbols represent the Modern Standard Arabic language and are designed to be culturally, socially, religiously, and linguistically acceptable. TAWASOL is based on the AAC Symbols Collection and is directed to nonverbal or minimally verbal individuals who require alternative communication solutions [25].

MAAN was developed using Xcode, which is an Apple integrated development environment (IDE) for developing iOS-based software. MAAN is available on iOS devices and is designed with a minimal number of screens. The screen designs are similar to traditional and commonly used messaging apps. Thus, in order to provide individuals with a consistent and effortless experience, TTS and STT features were implemented in the app using Xcode libraries. Additionally, 478 images from TAWASOL symbols were divided into 25 categories and added to MAAN. The essence of the category was to reduce the search time of required symbols. Participants can click on a certain category and then choose a TAWASOL symbol. The textual description of the selected TAWASOL symbol will be displayed in both Arabic and English. The user can specify the language of the symbol textual description, which is another advantage of MAAN over other apps. When previewing the received symbol, the user can listen to the TAWASOL symbol textual description. A challenging aspect related to designing for adults with ASD is the choice of the interface color theme [26,27]. MAAN aims to address this issue by allowing users to choose their suitable color theme. The MAAN app flow can be divided into the following 2 sections: administrative and features. Figure 1 shows screenshots of the app’s administrative flow, and Figure 2 shows screenshots of the app’s features.

Figure 1. Screenshots of the MAAN app’s administrative section.

In the administrative section, the app starts with a login screen where the participants can login using an assigned username and password, and remain logged in until they logout manually (Figure 1A). The participants can select whom to chat with and
start their conversation (Figure 1B). The setting screen provides the participants with the ability to view their profile, and change the app theme color and the incoming and outgoing messaging bubble color (Figure 1C and Figure 1D). Moreover, the participants can change the language of the app (Arabic or English) and logout from the app (Figure 1C).

The participants can use the features of the app as described in Figure 2. An accessory button in the chat window provides the participants with the choice to select the STT or TAWASOL symbols screen (Figure 2A and Figure 2B). On clicking the mic icon, they can record their voice and send it as a voice note to the person they are chatting with (Figure 2C). On selecting STT, the participants proceed to another screen where they can chat, and the chat will be directly transcribed into text to be sent (Figure 2D). On selecting TAWASOL symbols, the participants can choose symbols from the 25 different categories, which will be generated as text to be sent (Figure 2E). The last feature, TTS, will be activated when the participants press on any text message. With this feature, the system will transcribe the text into voice and read it.

Figure 2. Screenshots of the MAAN app’s feature display screens.
Testing and Analysis Tools

Ethical approval was obtained from the Research Board of Qatar Biomedical Research Institute. Participation was entirely voluntary, and each participant was sent an approved informed consent form that contained all study details. The following 4 types of consent forms were sent based on participant category: (1) participant without ASD (including caregiver or parent) consent form, (2) specialist consent form, (3) verbal participant with ASD consent form, and (4) nonverbal participant with ASD consent form. Each participant was aware of the requirement to provide an informed decision regarding participation in the study, as well as the right to withdraw from the study without justification or penalties. Moreover, participants were assured about the confidentiality and security of the data collected. The user evaluation study was conducted in a real-life setting, and the participants were recruited through snowballing techniques [28]. As of November 2020, a total of 7 adults with ASD, aged between 18 and 30 years, were recruited from Qatar and Lebanon. For each participant, parents, caregivers, or specialists were recruited for using the app.

Quantitative and qualitative measures were collected in this research. In terms of quantitative measures, this study employed 3 questionnaires and collected the interaction log data through the app. The prestudy questionnaire was used to collect the participants’ demographic data, and the SELSA-S [29,30] questionnaire was used to determine social and family loneliness for participants prior to the study and after completing the study. The uMARS [31] questionnaire was used to measure the app’s quality after completing the study. Interaction log data were also collected through Google Firebase. This included the number of times the app was accessed, and how each feature was used.

Quantitative data were complemented with qualitative data collected through the poststudy interview. The aim of the interview was to produce contextual real-world knowledge about the behaviors and social structures of the participants, and their experiences with the MAAN app in their daily lives. The interview was conducted with parents or specialists of the participants and was regarding social and communication skills development.

Quantitative Analysis

The first questionnaire used was the SELSA-S questionnaire. The full version of the SELSA questionnaire is a 37-item self-report measure of romantic, social, and family loneliness. It was administered to young adults with ASD prestudy and poststudy [29,30]. Moreover, the Program for Education and Enrichment of Relational Skills mainly used it in most of their studies [21] to track the social and loneliness aspects of their participants. A study showed that the SELSA-S questionnaire, which consists of 15 items, is a psychometrically reliable and valid alternative to the full version of SELSA, and requires less time for participants [32]. There are 5 items referring to the family loneliness subscale, 6 items referring to the romantic subscale, and 4 items referring to social loneliness [33]. All items are answered on a 7-point Likert scale. This study used only the social and family subscales, since emotional change was not in the interest of this study. The SELSA-S questionnaire was used prestudy and poststudy, and the results were compared to find the rate of change.

The Wilcoxon signed-rank test is a nonparametric test that is particularly suitable for examining the difference between pretest and posttest measures in a small sample [34], and it was used in this study. Statistical analysis was performed using IBM SPSS software (IBM Corp), and a P value <.05 was considered statistically significant.

The second questionnaire was the uMARS questionnaire that was coded following the scoring guides provided by the questionnaire developers [22]. The questionnaire consists of the following 3 sections: app quality, app subjective quality, and perceived impact. Moreover, the rating is scored out of 5, based on the scoring criteria given by the questionnaire developers. This questionnaire was used poststudy to evaluate the app’s quality. Descriptive statistical analysis of the uMARS rating was used to gain insights into the usability experience of the participants with ASD. Many studies have used the uMARS questionnaire to evaluate the usability of developed apps [35-37].

Qualitative Analysis

To analyze the interviews, open and axial coding phases from the grounded theory were used. Grounded theory is a systematic methodology extensively used in qualitative research. This approach aims to generate a substantive theory that links the investigated data to reality [38]. After conducting the poststudy interview with the participants, the interview was transcribed by the first author (MIFH). Open coding was then used to generate initial concepts from the data. This was followed by axial coding to establish connections between different concepts and categories [39].

Prestudy and poststudy interviews were conducted. The prestudy interview was conducted with 5 experts, 2 of whom worked in special educational programs (specialist and educational and ABA/applied verbal behavior consultant, and Information and Communications Technology access expert in educational programs). The other 3 worked in directing centers for adults with ASD (director of the Rehabilitation and Inclusion Office, manager of the Severe Difficulties Department, and president of the Disability Association). The interviews were conducted for the purpose of obtaining insights into the current situation and existing technologies (state-of-the-art), and obtaining feedback on the study design. The poststudy interview was conducted with all participants’ parents or caregivers to gather information about the context of use. The interview included questions about the overall app experience in terms of usability, functionality, and challenges encountered while using the app, as well as suggestions for further enhancement.

Results

Demographic Information

Among the 7 adults with ASD, 5 contacted their parents through the app, 1 contacted both a parent and specialist (ABA specialist and teacher with 25 years of experience), and 1 contacted only a specialist (speech and language therapist with 5 years of experience) (Table 1). Moreover, there was diversity in
educational background among the adults with ASD. This diversity in educational background was related to the severity of their conditions. Most of the parents and caregivers interacted with the participants for less than 6 hours per day. Based on the information gathered from the participants’ parents, 2 participants were diagnosed with high-functioning ASD (participants 3 and 4), 3 were diagnosed with medium-functioning ASD (participants 5, 6, and 7), and 2 were diagnosed with low-functioning ASD and up to severe ASD (participants 1 and 2). Participant information is presented in Table 2.

Table 1. Demographic information.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Adults with autism spectrum disorder (N=7), n</th>
<th>Parents (N=6), n</th>
<th>Specialists (N=2), n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Male</td>
<td>7</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-21</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>21-26</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>&gt;26</td>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 2. Characteristics of the participants with autism spectrum disorder.

<table>
<thead>
<tr>
<th>Characteristic described by a parent or specialist</th>
<th>Participant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completely nonverbal: Cannot speak but understands spoken conversation</td>
<td>Participants 1, 2, 3, and 5</td>
</tr>
<tr>
<td>Verbal: Can greet</td>
<td>Participant 4</td>
</tr>
<tr>
<td>Nonverbal: Cannot speak a lot and only greets, but understands spoken conversation</td>
<td>Participants 6 and 7</td>
</tr>
</tbody>
</table>

Participant Engagement

The user study evaluation lasted 16 weeks. However, not all participants completed this duration due to this project’s time limitation. Three participants completed the 16-week duration, whereas 2 participants used the app for 13 weeks, 1 participant used it for 11 weeks, and 1 participant used it for 10 weeks. The full duration of the study was tailored to each participant according to their enrolled duration. In this study, the duration was defined as the period from the start day after enrollment and training to the day when the poststudy interview was conducted. Participants’ daily activities were monitored through interaction log data. Table 3 represents a summary of the average number of days the app was accessed by each participant per week. Although it was difficult to determine who initiated a conversation, since it could last or hold for a while, initiation was considered when the participant began the chat each day. Based on responses in the poststudy interview, 2 participants initiated some of the conversations, especially when waking up in the morning, needing something to eat, or needing to go out (requesting a need); 2 participants partially initiated conversations with some encouragement from their parents or caregivers; and 3 participants did not initiate any conversations.

Table 3. Average number of days per week the app was accessed by each participant.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Value (days/week)a, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.00 (1.826)</td>
</tr>
<tr>
<td>2</td>
<td>2.37 (1.821)</td>
</tr>
<tr>
<td>3</td>
<td>2.87 (1.893)</td>
</tr>
<tr>
<td>4</td>
<td>4.62 (1.981)</td>
</tr>
<tr>
<td>5</td>
<td>1.69 (1.109)</td>
</tr>
<tr>
<td>6</td>
<td>3.36 (2.203)</td>
</tr>
<tr>
<td>7</td>
<td>2.10 (1.287)</td>
</tr>
</tbody>
</table>

aThe overall maximum and minimum values were 5 and 1.69 days/week, respectively, and the overall mean value was 3.14 (SD 1.25) days/week.

Based on the collected data, diversity in participant engagement was noticed. The participants were categorized into the following 2 groups: participants who used the app for 16 weeks and participants who used the app for less than 16 weeks. Figure 3 shows the average participant engagement for the 2 groups.
Figure 3. Participant engagement in the app.

Testing and Analysis Tools

Quantitative Analysis

In this study, the social and family loneliness subscales were considered. It was evident from the collected data that the family loneliness subscale did not change from prestudy to poststudy. However, the social loneliness subscale showed a difference in all participants (Figure 4).

The use of each app feature differed from one participant to another. However, it was evident that when the app features were frequently used, the SELSA-S rate of change tended to be higher (Figure 5). For example, participant 1 used TTS more often than all other participants, and thus, the SELSA-S rate of change of this participant was the highest. Participant 7 used the app features less frequently and did not get engaged with the app when compared with the other participants. The SELSA-S rate of change for this participant was low.

Figure 4. Prestudy and poststudy SELSA-S questionnaire social scale results. SELSA-S: short version of the Social and Emotional Loneliness Scale for Adults.
Regarding uMARS results, first, the app quality mean score was calculated by finding the mean of the ratings for the subsections engagement, functionality, esthetics, and information, and then calculating the average of the 4 means. The mean value was 3.62 (SD 0.65). Second, app subjective quality was rated based on the mean score of the related questions, which was 3.72 (SD 1.12). Third, the questionnaire’s perceived impact section consists of 6 questions that rate the impact of the app on the participants’ knowledge, attitude, and intention regarding the targeted health behavior. Moreover, the rating was based on the mean score of the related questions, which was 3.54 (SD 0.98).

Point-biserial correlation was employed to answer the first hypothesis. The number of times each app feature was used was considered a dichotomous variable, since it is either clicked or not [40,41]. Therefore, point-biserial correlation, which is a special case of Pearson correlation, was used to measure the strength of the association between the SELSA-S rate of change (post-pre) and each app feature. The strength of the correlation was assessed as follows: weak correlation if 0.1<|$r|$<0.3, moderate correlation if 0.3<|$r|$<0.5 and strong correlation if 0.5<$|r|$. The results showed that the SELSA-S rate of change was strongly related to the TTS feature ($r$=0.708, $P$=.38) and to the TAWASOL symbols feature ($r$=-0.917, $P$=.002). On the other hand, the SELSA-S rate of change was moderately related to the STT feature ($r$=0.428, $P$=.17). Therefore, usage of TTS, STT, and CS (TAWASOL symbols) can enhance the social and communication skills of adults with ASD. For the second hypothesis, the Wilcoxon signed-rank test was used to determine the difference between the prestudy and poststudy results in the social loneliness scale was significant. It indicated that poststudy SELSA-S results were statistically significantly higher than prestudy SELSA-S results ($z$=-2.047; $P$=.04). Thus, the null hypothesis was rejected. Therefore, usage of the app can reduce social loneliness among adults with ASD.

**Qualitative Analysis**

All the interviewees in the prestudy interviews agreed on the importance of the MAAN app as a unique tool that could be of great use. They stated that this innovative app would assist centers, parents, and caregivers in staying connected with adults with ASD. They also believed that MAAN has the potential to enhance the social and communication skills of adults with ASD. Moreover, they asserted that such an app could support the inclusion of adults with ASD into social, educational, and work settings, ultimately encouraging adults with ASD to use more social-based apps.

Overall, 6 parents and 1 specialist were interviewed in the poststudy interview. The interview focused on their experiences with the app, the challenges faced, and the recommended modifications. The open and axial coding phases of the grounded theory [39] were used to analyze the transcribed interview. The analysis highlighted the importance of the app in enabling communication between adults with ASD and other individuals. Most interviewees discussed the relevance of the app and its novelty. Five categories emerged from the analysis, which were app support, feature relevance, user interface design, overall feedback, and recommendations. Each category had subcategories and samples of excerpts from the transcribed data describing the experiences of the participants with ASD while using the app (Table 4).
<table>
<thead>
<tr>
<th>Category and subcategory</th>
<th>Description</th>
<th>Example quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>App support</strong></td>
<td>Technical and customized support to access the needed information without problems or errors</td>
<td></td>
</tr>
<tr>
<td>Learnability</td>
<td>Interactive app features, which aid participants with ASD⁴ to learn quickly</td>
<td><em>He was able to understand how to use it after I trained him...</em> [Parent #4]</td>
</tr>
<tr>
<td>Efficiency (accessibility)</td>
<td>Accessing the features of the app (easy, medium, or hard)</td>
<td><em>He accepted the application very smoothly and enjoyed using it...</em> [Parent #4]</td>
</tr>
<tr>
<td>Technical issues</td>
<td>Bugs or related technical issues</td>
<td><em>He tried to use recording, but it was a little hard for him to use it...</em> [Parent #5]</td>
</tr>
<tr>
<td><strong>Feature relevance</strong></td>
<td>Feedback of the participants on the app features</td>
<td></td>
</tr>
<tr>
<td>Text-to-speech</td>
<td>Feedback on the text-to-speech feature</td>
<td><em>...it was an amazing experience for him to listen to the text, while this feature is not available in any other platform even WhatsApp...</em> [Parent #1]</td>
</tr>
<tr>
<td>Speech-to-text</td>
<td>Feedback on the speech-to-text feature</td>
<td><em>...Speech to text is a very good feature I would say especially for autistic case...</em> [Parent #6]</td>
</tr>
<tr>
<td>Arabic issues</td>
<td>Participants talk about Arabic issues related to the app</td>
<td><em>...The Arabic language was a little annoying since text-to-speech and speech-to-text transcriptions were robotic and need more advancements...</em> [Parent #2]</td>
</tr>
<tr>
<td>TAWASOL symbols</td>
<td>Feedback on TAWASOL symbols</td>
<td><em>...For him, both TAWASOL symbols and voice messaging are the main features that helped him...</em> [Parent #2]</td>
</tr>
<tr>
<td><strong>User interface design</strong></td>
<td>Visual appearance of the app, such as the arrangement of content, color schemes, icons, and font sizes</td>
<td></td>
</tr>
<tr>
<td>Color</td>
<td>Feedback on the color changing tool of the app</td>
<td><em>...He didn’t change the application color nor the chat bubble since he doesn’t have any problem with colors, but other autistic individuals can use this feature to help them using the application more by changing the color with what is suitable for them...</em> [Parent #2]</td>
</tr>
<tr>
<td>Design consistency</td>
<td>The participant likes texting and using the app like other messaging apps</td>
<td><em>...He knows [referring to the participants with ASD] how to type and use it in other application, then he liked the chatting (Texting) part more...</em> [Parent #1]</td>
</tr>
<tr>
<td><strong>Overall feedback</strong></td>
<td>Feedback about the app in general regarding the innovative idea and the need of it</td>
<td></td>
</tr>
<tr>
<td>Positive feedback</td>
<td>Positive feedback on the app</td>
<td>*...Really good application for chatting like I'm very happy about the way it sends messages in different format text-to-speech, speech-to-text, and Tawasol symbols, so I was introducing it to him as a chatting app itself and he liked it [Caregiver #1]</td>
</tr>
<tr>
<td>Novelty</td>
<td>Feedback on the novelty of the app compared to existing messaging apps</td>
<td><em>...I’m just saying because it’s a very good with its unique features what I really found very interesting since I can’t find these features in other similar applications such as WhatsApp...</em> [Parent #1]</td>
</tr>
<tr>
<td><strong>Recommendations</strong></td>
<td>Suggestions of parents and specialists on how to improve the app</td>
<td></td>
</tr>
<tr>
<td>Serious games</td>
<td>Suggestions to enhance the app with game elements to teach specific skills and knowledge</td>
<td><em>...I would like to suggest including games where autistic adults could learn the TAWASOL symbols and be able to easily construct sentences...</em> [Parent #3]</td>
</tr>
<tr>
<td>Privacy protection</td>
<td>Discussing the next version based on a friend request feature in order to protect users from strangers</td>
<td><em>...Caregivers should be able to accept friends, and this is because we need to protect them from unknown people...</em> [Parent #3]</td>
</tr>
</tbody>
</table>

⁴ ASD: autism spectrum disorder.
Discussion

Summary

In this study, MAAN, a mobile messaging app, was developed and evaluated in 7 adults with ASD over 10 to 16 weeks. A pre-post study was also conducted with experts and parents on the interface and functionality design of the app. The app has additional features (TTS, STT, and CS) when compared with existing messaging apps. MAAN is designed to support adults with ASD when communicating with other individuals via text messaging in both Arabic and English. It offers the ability for adults with ASD to read or listen to text messages and then reply. The results supported the hypotheses and are encouraging for further work in the future.

Comparison With Related Work

Very few messaging apps that support social and communication skills in individuals with ASD are discussed in the literature. TalkingBoogies [42], an app introduced in 2020 by a research team in Korea, aimed to actively assist caregivers when developing AAC-led communication with children having ASD. It comprises the following 2 iOS apps: TalkingBoogies-AAC for child caregiver communication and TalkingBoogies-Coach for caregiver collaboration. TalkingBoogies-AAC has several features, such as the use of Ewha [43], an AAC symbol system, and TTS. The team evaluated the developed app in 4 children with ASD and 3 teachers with at least 6 months of experience with AAC from a local special education school. The study concluded that such an app can prompt the learnability of children when constructing sentences. Prior to this study, De Leo et al [44] built and evaluated a Windows-based mobile app named “PixTalk.” The PixTalk system employs the PECS intervention and is made up of the following 2 modules: PixTalk smartphones, which enables children with ASD to search and select images in order to express their needs and feelings, and PixTalk website, which allows caregivers and teachers to add different images to a child’s smartphone. By conducting a case study involving 3 children with ASD and their teachers, this research asserted the importance of computer-assisted instructions as an intervention to motivate and engage learners with ASD. Proloquo2Go [45] and Proloquo4Text [46] are popular communication apps that use symbols and the TTS feature. Both apps are commercially available and widely used in school settings where English is the main instructional language. Nowadays, virtual communication platforms, such as social media networks and messaging apps, are becoming more integrated in most people’s lives [47]. However, from an accessibility viewpoint, the use of these platforms is neither practical nor accessible for the majority of individuals with ASD and especially for the nonverbal population. Some of the existing work involves AACrobat [48]. The authors designed this mobile app to help neuromuscular disease users communicate with others using their eyes. Although AACrobat uses the TTS feature, this app is not helpful for individuals with ASD since most of them cannot fix their eyes in one direction to type from a keyboard or do not have the ability to generate a sentence. Currently, the most popular communication platform is WhatsApp [47], which is mainly used as a messaging app. Despite all the important updates and features that WhatsApp provides, it is still not accessible for individuals with ASD, as most of the nonverbal population struggles to read, write, or communicate using basic texting features. To the best of our knowledge, previous studies have attempted to support adults with ASD in constructing sentences for education and learning purposes, but no study has evaluated TTS, STT, and CS features, as well as social and family loneliness in adults with ASD. This paper presents the development and evaluation of the new messaging app MAAN, which means “together” in Arabic. MAAN employs TTS, STT, and CS to enhance the social and communication skills of adults with ASD. This is the first study to consider these features in a mobile messaging app designed to support communication among adults with ASD, both in English and Arabic. The results showed that MAAN can enhance social and communication skills, especially through distance messaging, and can ease communication with peers through its special features (TTS, STT, and CS). Hence, this study showed the effectiveness of these features in a messaging app.

Principal Results

The 5 experts interviewed in the prestudy phase emphasized the importance of MAAN as a unique tool in supporting adults with ASD. During the interview, they discussed the app’s novelty and suggested several modifications, such as categorizing TAWASOL symbols and fixing the app’s design. Moreover, they emphasized the support of the app for the inclusion of adults with ASD into social, educational, and work settings, ultimately encouraging adults with ASD to use more social-based apps. On the other hand, the poststudy interview with parents suggested that the app provides 3 features directed to verbal and nonverbal adults with ASD, and the TTS feature was the most preferable feature for both ASD and non-ASD participants. Using MAAN helped the participants with ASD interact more with their parents, especially when they were not near them. Besides, the participants with ASD were attached to the app, where they voiced their needs by sending a message to their parents. Support for the Arabic language in all 3 features made this app very appealing to the participants, considering that there is a lack of ACC apps that support the Arabic language. Moreover, the interviewees highlighted the importance of MAAN in enhancing the social and communication skills of different users, such as individuals with dyslexia. From the participants’ engagements with the app and the SELSA-S questionnaire results, it can be deduced that more use of the app features was associated with higher SELSA-S questionnaire scores. Moreover, this was confirmed in the poststudy interviews where the interviewees highlighted the importance of the app features and their ability to increase the attention of adults with ASD, which can positively improve social loneliness. This development is noticeable with the positive engagement and enthusiasm that the participants with ASD exhibited when messaging their parents using the app features or when constructing a sentence by using TAWASOL symbols.

Limitations

Despite this study’s contributions, some limitations, including the sample size, operating system platform, and choice of features, were noted. Due to the COVID-19 pandemic, most
adults with ASD were at home with their families and thus could not be easily reached. This led to a rather small sample. Another factor that contributed to the small sample was the absence of focused institutions for adults with ASD, since most of the current institutions serve children with ASD. A larger sample size could give more insights into the effects of the app and could increase the statistical power. Further, the message app was only available on the iOS platform. This availability limitation was due to the timeframe for developing the app. Lastly, other messaging app features, such as image messaging, video messaging, and voice and video calls, were excluded. The reason is to prompt adults with ASD to use AAC-based features rather than video or voice calls, which, in turn, can support their social and communication skills.

Conclusion
The novelty of MAAN as a communication and social intervention app is its potential to support communication skills and social loneliness in adults with ASD. More importantly, it was possible to achieve this due to the inclusion of experts in the design and development of the app. Additionally, the poststudy evaluation by parents and specialists identified the uniqueness of the app, and how it could be enhanced and extended to other populations who also exhibit social and communication deficits, such as people with dyslexia. Future studies can consider a larger number of participants with ASD to replicate the findings and can extend this study to other clinical populations with social and communication deficits.

Acknowledgments
We would like to express our gratitude to Hamad Bin Khalifah University (HBKU). We also gratefully acknowledge all the participants who took part in the study. Their dedication and commitment have had an enormous impact on the quality of this research. We would like to thank the administration of Step by Step Center (for Special Educational Needs) and Best Buddies Center in Qatar, and the Autism Learning Institute for Applied Behaviour Analysis (ALI for ABA Center; Dr Chafica Mansour Gharbieh) in Lebanon for allowing their teaching staff, specialists, and users to take part in the study.

Conflicts of Interest
None declared.

References


47. SPECTRM. 2020. URL: https://tinyurl.com/yckuycf3 [accessed 2021-02-24]


Abbreviations

AAC: augmentative and alternative communication
ABA: applied behavioral analysis
ASD: autism spectrum disorder
CS: communication symbols
PECS: picture exchange communication system
SELSA-S: short version of the Social and Emotional Loneliness Scale for Adults
STT: speech-to-text
TTS: text-to-speech
uMARS: Mobile Application Rating Scale: user version

Edited by G Eysenbach; submitted 24.08.21; peer-reviewed by 신, MDG Pimentel; comments to author 03.10.21; revised version received 08.11.21; accepted 17.11.21; published 07.12.21.

Please cite as:
Hijab MIF, Al-Thani D, Banire B
A Multimodal Messaging App (MAAN) for Adults With Autism Spectrum Disorder: Mixed Methods Evaluation Study
JMIR Form Res 2021;5(12):e33123
URL: https://formative.jmir.org/2021/12/e33123
doi: 10.2196/33123
PMID:34878998

Juan Espinoza¹,², MD; Abu Taher Sikder³, BA; James Dickhoner²,³, MD; Thomas Lee²,⁴, MD

¹Department of Pediatrics, Children’s Hospital Los Angeles, Los Angeles, CA, United States
²Keck School of Medicine, University of Southern California, Los Angeles, CA, United States
³Innovation Studio, Children’s Hospital Los Angeles, Los Angeles, CA, United States
⁴Department of Surgery, Children’s Hospital Los Angeles, Los Angeles, CA, United States

Corresponding Author:
Juan Espinoza, MD
Department of Pediatrics
Children’s Hospital Los Angeles
4650 Sunset Blvd
Los Angeles, CA, 90006
United States
Phone: 1 3233612721
Email: jespinoza@chla.usc.edu

Abstract

Background: Health care databases contain a wealth of information that can be used to develop programs and mature health care systems. There is concern that the sensitive nature of health data (e.g., ethnicity, reproductive health, sexually transmitted infections, and lifestyle information) can have significant impact on individuals if misused, particularly among vulnerable and marginalized populations. As academic institutions, nongovernmental organizations, and international agencies begin to collaborate with low- and middle-income countries to develop and deploy health information technology (HIT), it is important to understand the technical and practical security implications of these initiatives.

Objective: Our aim is to develop a conceptual framework for risk stratification of global health data partnerships and HIT projects. In addition to identifying key conceptual domains, we map each domain to a variety of publicly available indices that could be used to inform a quantitative model.

Methods: We conducted an overview of the literature to identify relevant publications, position statements, white papers, and reports. The research team reviewed all sources and used the framework method and conceptual framework analysis to name and categorize key concepts, integrate them into domains, and synthesize them into an overarching conceptual framework. Once key domains were identified, public international data sources were searched for relevant structured indices to generate quantitative counterparts.

Results: We identified 5 key domains to inform our conceptual framework: State of HIT, Economics of Health Care, Demographics and Equity, Societal Freedom and Safety, and Partnership and Trust. Each of these domains was mapped to a number of structured indices.

Conclusions: There is a complex relationship among the legal, economic, and social domains of health care, which affects the state of HIT in low- and middle-income countries and associated data security risks. The strength of partnership and trust among collaborating organizations is an important moderating factor. Additional work is needed to formalize the assessment of partnership and trust and to develop a quantitative model of the conceptual framework that can help support organizational decision-making.

(JMIR Form Res 2021;5(12):e25833) doi:10.2196/25833

KEYWORDS

health information technology; low- and middle-income countries; low income; conceptual framework analysis; framework method; data security; decision-making; database; information use; misuse; global health; security
Introduction

Background

Health information technology (HIT) refers to electronic health records (EHRs), patient portals and other software platforms, public health databases, hardware devices, and technology systems, which contain a wealth of information used for patient care and resource allocation. According to a 2016 World Health Organization (WHO) global survey on eHealth, the adoption of EHR systems had increased by 46% in the previous 5 years [1]. Increasing numbers of low- and middle-income countries (LMICs) are implementing HIT as they continue developing their health care infrastructure [2,3]. In many cases, these systems are being implemented in collaboration with foreign academic institutions, health care systems, and nonprofit or research organizations [4,5]. However, there are unique organizational, technical, functional, educational, and ethical challenges that require meticulous consideration—especially with respect to their security implications [6].

In many LMICs, legal, regulatory, and technical frameworks around HIT are undeveloped [6]. Furthermore, the sensitive nature of health data (e.g., ethnicity, reproductive health, sexually transmitted infections, and lifestyle information) can have significant impact on individuals if misused [7]. The combination of developing legal frameworks and decreased ability of public institutions to protect individuals may create a particularly vulnerable environment for HIT and health data. A framework to understand and stratify the risk associated with HIT may be beneficial to organizations engaging in global health partnerships that generate significant amounts of health data.

Our research team has been focused on international clinical and research partnerships in Armenia. As we have engaged in the process of designing and deploying a safe, scalable health data platform in Armenia, other countries in the region have expressed interest in implementing similar systems. This geopolitical region is home to the Commonwealth of Independent States (CIS), an intergovernmental organization of 11 post-Soviet countries: Armenia, Azerbaijan, Belarus, Kazakhstan, Kyrgyzstan, Moldova, Russia, Tajikistan, Turkmenistan, Ukraine, and Uzbekistan (Figure 1) [8]. These countries inherited the Soviet Union’s Semashko health system and many of its flaws in transitioning to a modern primary care model, including premature mortality, variable quality of care, poor noncommunicable disease management, and high out-of-pocket payments [9-11]. Rising social and economic challenges such as inequality and the cost of funding public health are also significant concerns [9]. To address these concerns, CIS member states have embarked on health care reform efforts to improve their health care systems, including deploying HIT [9,12].

Figure 1. A map of Commonwealth of Independent States countries located in Eastern Europe and Central Asia. The color gradient shows 2019 gross domestic product per capita in US dollars. The population of the country (in millions) is noted next to the name. GDP: gross domestic product.

Need for a Framework

As our organization considers new partnerships among other CIS members, the need for a framework to understand potential HIT security risks has become more pressing. Given the lack of an existing risk stratification framework to assess and consider the security vulnerabilities with implementing HIT in LMICs, we decided to create our own. In this paper we present a conceptual approach to developing such a framework and propose a variety of indices that could be leveraged to inform each subdomain. Finally, we outline our proposed next steps to
seek consensus and finalize quantitative and qualitative versions of the framework.

**Methods**

**Literature Review**

We reviewed the literature to identify relevant publications by searching PubMed, Ovid MEDLINE, Google Scholar, and Google using the search terms presented in Textbox 1. All articles were identified as relevant by at least 2 authors (JE, ATS, or JD). References from relevant articles were also reviewed. Original research, reviews, editorials and commentaries, position statements, white papers, and industry and nongovernmental organization reports were included. Finally, the websites of international agencies such as the WHO, the United Nations, and the World Bank were reviewed for relevant data sources and publications.

Textbox 1. Literature search terms by concept type.

<table>
<thead>
<tr>
<th>Search terms</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Geographic terms</td>
<td>Armenia</td>
</tr>
<tr>
<td></td>
<td>Europe</td>
</tr>
<tr>
<td></td>
<td>Asia</td>
</tr>
<tr>
<td></td>
<td>Soviet</td>
</tr>
<tr>
<td></td>
<td>Commonwealth of Independent States (CIS)</td>
</tr>
<tr>
<td></td>
<td>Low- and middle-income countries (LMIC)</td>
</tr>
<tr>
<td></td>
<td>Developing countries</td>
</tr>
<tr>
<td></td>
<td>Global health</td>
</tr>
<tr>
<td></td>
<td>International</td>
</tr>
<tr>
<td>Technology terms</td>
<td>Data</td>
</tr>
<tr>
<td></td>
<td>Health data</td>
</tr>
<tr>
<td></td>
<td>Patient data</td>
</tr>
<tr>
<td></td>
<td>Information</td>
</tr>
<tr>
<td></td>
<td>Health information</td>
</tr>
<tr>
<td></td>
<td>Patient information</td>
</tr>
<tr>
<td></td>
<td>Electronic health records (EHR)</td>
</tr>
<tr>
<td></td>
<td>Electronic medical records (EMR)</td>
</tr>
<tr>
<td></td>
<td>Health information technology (HIT)</td>
</tr>
<tr>
<td></td>
<td>Personal health records</td>
</tr>
<tr>
<td></td>
<td>Database</td>
</tr>
<tr>
<td>Theme terms</td>
<td>Security</td>
</tr>
<tr>
<td></td>
<td>Risk</td>
</tr>
<tr>
<td></td>
<td>Partnership</td>
</tr>
<tr>
<td></td>
<td>Breach</td>
</tr>
<tr>
<td></td>
<td>Trust</td>
</tr>
<tr>
<td></td>
<td>Safety</td>
</tr>
<tr>
<td></td>
<td>Health system</td>
</tr>
<tr>
<td></td>
<td>Health policy</td>
</tr>
<tr>
<td></td>
<td>Legal</td>
</tr>
</tbody>
</table>
Conceptual Framework Development

A conceptual framework is a “network of...interlinked concepts that allow for a comprehensive understanding of a phenomenon or phenomena” [13]. Our conceptual framework development process borrowed iterative techniques from both the framework method [14] and conceptual framework analysis [13]. First, all literature sources were extensively reviewed for key concepts. By categorizing and recategorizing these key concepts iteratively as more literature was reviewed, the team conceptualized them into domains that were then synthesized to develop an overarching conceptual framework using the coding methodology outlined by Gale et al [14]. Once key domains were identified, public international data sources were searched for relevant structured indices to generate a quantitative counterpart (Figure 2).

**Figure 2.** A visual representation of our approach to developing our conceptual framework, from literature review to refinement. NGO: nongovernmental organization.

Expert Input and Refinement

Our conceptual framework, key domains, and indices were shared formally and informally through email, conversations, and presentations with international colleagues in health care, information technology, informatics, health policy, and international relations. Feedback was collected and an iterative approach was used to refine the model and the indices included.

Results

Key Domains and Indices

Our analysis resulted in an inventory of concepts that aggregated into 5 key domains that make up our conceptual framework for assessing health data security considerations in global health partnerships:

1. State of HIT
2. Economics of Health Care
3. Demographics and Equity
4. Societal Freedom and Safety
5. Partnership and Trust

For each of the first 4 domains, we identified publicly available indices that could be used to provide qualitative or quantitative insights. For the fifth domain, Partnership and Trust, we created a self-assessment tool that would allow an organization’s members to take into consideration the specifics of their project, their experience on the ground, and their in-country partners.

State of HIT

**Overview**

HIT refers to all the electronic systems used by care providers, public health workers, patients, researchers, and others to manage health information [15]. It includes EHRs, other software platforms, hardware devices, and technology systems [16]. HIT systems are a critical component of several aspects of health care, including care delivery, billing, medicolegal liability, research, and health policy [17,18]. HIT use is influenced by factors such as cost, technical feasibility, and regulation [3].

The State of HIT domain evaluates a health system’s maturity in terms of its logistic, technological, and regulatory progress within a country. The legal structure and regulatory framework around HIT and medical ethics directly influence the viability and progress of HIT [19]. As suggested by Luna et al [19], overcoming legal and ethical challenges, interoperability issues, and technical security vulnerabilities greatly affects the implementation of HIT.

Rationale for Indices

Multimedia Appendix 1 [6] presents the proposed indices of this domain categorized into 3 broad subdomains:

1. EHR deployment considers the presence of a national EHR system’s existence, regulation, and location. National EHR systems require a high level of commitment as well as technical and financial resources [3,20]. This implies a certain level of investment, commitment, and organizational sophistication and provides insight into the groundwork
laid by the country that affects the long-term viability and integrity of HIT and health data [20,21]. Were at al [20] have commented on the unsustainability of expensive implementations and the need for sustainability to be built into EHR implementations, underscoring the demand that EHRs place on technical and financial resources. This subdomain also considers EHR implementation location (in primary, secondary, or tertiary facilities), which gives an understanding of the level of penetration of the technology in the market. Overall, EHRs are often the first, most expensive, and most expensive component of HIT; as such, they serve as reasonable proxies for HIT overall, including technical and logistical risks with implementing new technologies, where other measures are not available. The indices in Multimedia Appendix 1 [6] were gathered from the Electronic Health Record section of the Atlas of eHealth Country Profiles compiled by the WHO as “Yes, No, Not applicable [6].”

2. eHealth Foundations refers to the national strategy, policies, and funding information for eHealth. These factors are critical to determining the success or failure of HIT and can provide valuable insights into risk stratification of a given environment [22]. These indices were gathered from then eHealth Foundations section of the Atlas of eHealth Country Profiles compiled by the WHO as “Yes, No, Not applicable [6].”

3. Legal Frameworks for eHealth refers to the policies, regulation, and jurisdiction governing the use, quality, sharing, and ownership of health data. Medicolegal concerns are universal, complex challenges of HIT that affect use and the experience of patients, providers, and other stakeholders [20,23]. The existence of legal solutions or mechanisms to address these concerns can provide insight into the preparedness of a country to mitigate HIT and data breaches as well as suggest the recourse that individuals have should a breach occur. For example, Palabindala et al [23] mention that Health Insurance Portability and Accountability Act compliance requires substantial legal, technical, and logistical efforts that ensure the establishment of appropriate measures for unfortunate events. These indices were gathered from the Legal Frameworks for eHealth section of the Atlas of eHealth Country Profiles compiled by the WHO as “Yes, No, Not applicable [6].”

Economics of Health Care

Overview

The Economics of Health Care domain aims to quantify a country’s overall investment in health care. This domain evaluates a country’s financial and resource investment in health care access, delivery, services, and technology. Government investment directly affects infrastructure, quality of care, and affordability for patients and providers, which in turn significantly influences the successful and sustainable implementation of HIT, especially in LMICs [19,24,25]. Without proper allocation of funding, resources, care standards, and cost-effectiveness for stakeholders, there is a greater chance of failure in terms of long-term success and scalability [20]. As discussed by Luna et al [19], financial and technical sustainability is an important element of HIT integration and therefore needs to be addressed before implementation.

Rationale for Indices

This domain can be further evaluated as 4 subdomains; the individual indices have been presented in Multimedia Appendix 2 [26-43]:

1. Health Care Expenditure considers gross domestic product (GDP), contextualized GDP for health care, and health care expenditure. These indices were gathered from data by the World Bank on GDP, GDP per capita, GDP growth percentage, health care spending percentage per GDP, and health care spending per capita in US dollars, US dollars per capita, or percentages depending on the metrics involved [26-31]. Indices such as recontextualized GDP values and health care spending per capita have been described by Naik et al [44] as macroeconomic determinants of health that ultimately influence HIT integration and management. Insufficient health care funding increases the chances of failure of HIT implementation and health data management [20]. Health expenditure and GDP spending on health care thus become potential proxies for risk assessment.

2. Health Care Structure considers both infrastructure and system indices such as hospital beds, doctors per capita, health care access, and the existence of a public health care system. These indices were gathered from data sets by the World Bank on doctors per capita and hospital beds, the Global Burden of Disease (GBD) index on health care access and quality, and the US Social Security Administration on private versus public health care systems as numbers (per 1000 people), indices (1-100), and binary values [32-35]. As these indices directly measure access and the availability of crucial health care resources, they provide insight into the economic and material context of health care. Lower scores for these indices may highlight a higher risk of not having adequate finances and resources for implementing and managing HIT systems [3]. Specifically, Akhlaq et al [3] identified infrastructure, finance, organization, and data management as key factors in the adoption and management of HIT.

3. Health Care Cost considers out-of-pocket fees and universal health coverage. These indices were gathered from the 2017 Global Monitoring Report by the World Bank on universal health care coverage as values (0-100) measuring affordability and data sets by the World Bank on out-of-pocket costs as a percentage of total universal coverage [36,37]. These indices are important because they provide insight into the patient-level microeconomic context. Cost-effectiveness and financial viability for patients directly affect access, use, and long-term potential of health care services and resources, including HIT [19]. A lack of affordable health care can create risks for the overall success of health HIT.

4. Health Care Quality is intended to evaluate overall health care system performance through process and outcome measures such as health performance index, infant and maternal mortality, life expectancy, immunization rates, and diarrheal disease rates. These indices were gathered from data sets by GBD collaborators, the WHO, the Central Intelligence Agency, and the GBD database as an index.
from 1 to 191, deaths per births, age of death, and percentage immunized [38-43]. The indices listed in Multimedia Appendix 2 [26-43] are frequently used in the literature and by international organizations to measure overall health care quality [25,45,46]. Measuring health care quality provides a lens through which to interpret the economic inputs of a health care system. Major discrepancies between economic inputs and health care quality outcomes may be a cause for concern because these may result from a variety of issues, including health care administration, system stability, and inadequate data collection and reporting. All of these would be factors that may affect health data security.

Demographics and Equity

Overview

The Demographics and Equity domain aims to describe the relevant population and possible disparities involving health care. Understanding how patients from different social, economic, ethnic, and cultural backgrounds experience HIT is important for any global health data partnership because these differences can drive care disparities [25,47]. Gathering together concepts of population demographics, social structure, and community development provides a starting point to gain necessary insights. These social and demographic variables help contextualize pragmatic concerns surrounding patient privacy, access, health discrepancies, and improper use of health data [20]. Increased digitization of health care has several advantages such as public health surveillance during COVID-19, but this same surveillance infrastructure has implications for civil liberties and governance that affect marginalized groups differently [48]. For example, the Social Science Research Council states that Black and Brown communities are subject to disproportionate police surveillance and may be unable to opt out of medical tracking and monitoring systems [48]. Being aware of these issues in the local context is an important component of responsible data stewardship.

Rationale for Indices

Multimedia Appendix 3 [49-61] presents the indices of this domain categorized into 3 broad subdomains:

1. Population metrics include information about the density and structure of the population. The indices were gathered from data by the World Bank on population age structure as percentages and from the United Nations on rural and urban population density in thousands [49-52]. These population metrics affect HIT in a variety of ways. For example, countries with larger populations may require more costly efforts to ameliorate data misuse [62,63]. In addition, differences in age structure and trends may affect the demand and risk of health services and data technology [64-66]. As noted by Knickman and Snell [64], the Baby Boom generation is expected to double by 2030 and will require substantially more health-related resources. The increasing financial demand and use of health data may affect the technical and logistical risks associated with HIT.

2. Social Structure includes information about wealth inequality, poverty, decentralization, and public trust. The indices were gathered from data by the World Bank, a policy paper by World Bank affiliates titled How Close Is Your Government to its People, and Edelman, a global communications firm, as either a percentage or score (1-100), as appropriate [53-56]. Wealth inequality and poverty data provide insights into the economic aspect of inequality. Decentralization has long been advocated by international development agencies to improve health system performance in LMICs [67]. A recent literature review showed limited empirical data to support this approach, but as a dominant theory in international development, it should still be considered [68]. A lack of public trust in government and health care systems can lead to poor patient compliance with public health guidance, delaying seeking care, and withholding of critical information from providers [69]. This can lead to incomplete or unreliable data. Low public trust can be a symptom of either systematic failures of health systems or breaches of trust at the interpersonal level, both of which should be taken into consideration when discussing data privacy and security.

3. Community Development includes information on human development, access to electricity and the internet, literacy rate, and social media penetration. The Human Development Index is a composite measure developed by the United Nations that quantifies the capability of an individual to live a long and healthy life and acquire resources for a basic standard of living as a value from 0 to 1 [57]. Other descriptors of community development that focus on the ability of a community to meaningfully leverage technology are also included. Data sets from the World Bank on internet subscribers, access to electricity, and literacy were gathered as a number (1-100) or percentage as appropriate [58-60]. Social media penetration was gathered from Statcounter Global Stats as a percentage [61]. Basic resources such as electricity and the internet are necessary to meaningfully interact with HIT. This is true at the level of health care facilities as well as at the individual level [2]. Settings with limited access to the internet and electricity may not be able to implement a wide variety of privacy and security tools such as two-factor authentication. Limited literacy can be a barrier to people’s ability to use and access technology and data, which may make them more vulnerable to exploitation [70,71]. This increases the human cost of inadequate privacy and security in health care. More specific concepts such as health and technology literacy may be relevant and should be explored further [70,71].

Societal Freedom and Safety

Overview

Societal freedom is the liberty of an individual to function in society without coercion; the Cato Institute defines this as “the dignity of an individual [72].” As an ever-present societal factor, it influences aspects of health care both directly and indirectly [20,24,47,73]. The Societal Freedom and Safety domain aims to quantify the absence of coercive societal constraints on individual freedom within a country as well as the robustness of civil and political liberties; it includes concepts such as liberty of expression, social organization, and lawfulness. Overall,
countries with more freedom (democratic) have more robust health care systems and lower mortality than countries with less freedom (autocratic) [74,75]. Factors such as the rule of law and the influence of civil society affect health and health outcomes [76-78]. Pinzon-Rondon et al [76] found that adherence to the rule of law is associated with a healthier population, higher life expectancy, and lower adverse health outcomes. These social parameters provide insights into how likely a malicious data breach might be; whether the threat to personal sensitive data might come from government or nongovernment actors; and how vulnerable individuals may be to adverse social, financial, and legal consequences in case of a breach of privacy with respect to their personal health data.

**Rationale for Indices**

Multimedia Appendix 4 [72,79] presents the indices of this domain along with an overall rank and score:

1. Personal Freedom includes information on the rule of law, safety and security, religious freedom, assembly and association, expression, and identity and relationship. These indices were gathered from the Human Freedom Index 2019 by the Cato Institute as an index value between 0 and 1 [72]. Personal freedom is important to consider for data security because it highlights societal challenges with data. For example, religious hostility and persecution, surveillance of expression and information, geopolitical concerns, and stigma toward historically marginalized groups may increase the risk associated with access to identifiable health information [7,80]. In a 2010 report by the London School of Economics, the authors stressed the importance of social context and appropriate safeguards for HIT implementation, given the vast cultural and environmental differences that can exist even within a country [80]. In addition, limitations in assembly and association may adversely influence public health initiatives and health policy [78]. These concepts can provide a more nuanced assessment of the data security risk.

2. Economic Freedom includes economic liberty, sound money, property rights, international trade, and regulation of financial institutions. These indices have been gathered from the Human Freedom Index 2019 by the Cato Institute as an index value between 0 and 1 [72]. Economic freedom provides insights into the financial opportunities for individuals and organizations. In societies with high economic freedom there may be additional economic incentives to develop robust HIT [81]. There may also be opportunities for private and public-private partnerships to enhance data security [82].

3. Global Freedom is a concept developed by Freedom House, a US-based nonprofit focused on promoting democracy, and published in their annual Freedom in the World Report since 1973 [79,83]. It is a quantitative and qualitative assessment of political rights and civil liberties in countries and territories around the world, represented as a weight score on a scale from –4 to 100. Evaluations of a country’s electoral process, political participation, government functioning, associational rights, rule of law, and personal autonomy make up the global freedom score. Freedom House also publishes other relevant indices such as the Internet Freedom Score and Democracy Score, but these cover a significantly smaller number of the world’s countries (65 countries and 29 countries, respectively) and thus may not be as helpful in creating a standard analytical approach.

**Partnership and Trust**

An increasing number of partnerships have been developing between high-income countries and LMICs to address the global burden of disease. The success of these projects requires strong partnerships, which involves establishing rapport [84]. Relationship building and trust have been shown to be critical in navigating pragmatic obstacles as well as cultural and logistical boundaries [85,86]. Wagner et al [85] highlight how local coordinators and hosting communities are vital for the execution of international projects, and therefore establishing relationships and promoting ongoing collaboration are imperative to the success of global health efforts.

The strength of partnerships may be a moderating factor for concerns around patient data misuse. Organizations should objectively evaluate their global health partnerships. However, there is limited literature on global health partnership assessment tools. Instruments such as the Partnership Assessment Toolkit do exist, but studies need to be conducted to better understand their uses, limitations, and effectiveness [87,88]. To our knowledge, there are no tools that specifically address health data concerns. Given that each partnership is unique and influenced by a number of factors, we propose that this domain should be a self-assessment completed by the collaborating organizations. Relevant questions to explore include details about the in-country partner, the length of time the partnership has existed, the scope of the partnership, sensitive personal data collection, and relevant data security expertise resources available to the partners. Textbox 2 presents a list of potential questions to include in a self-assessment.

<table>
<thead>
<tr>
<th>Questions to include in a self-assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>• How long have your organizations worked together?</td>
</tr>
<tr>
<td>• How long has your partner been active in-country?</td>
</tr>
<tr>
<td>• How much experience does your partner have with health data?</td>
</tr>
<tr>
<td>• Is your local partner in good standing in-country?</td>
</tr>
<tr>
<td>• What is your organization’s reputation in-country?</td>
</tr>
<tr>
<td>• Is an official government entity with oversight over health, health care, or data involved in your project? If not, should they be?</td>
</tr>
<tr>
<td>• Are there known examples of health data misuse in the country?</td>
</tr>
<tr>
<td>• In your partnership, who owns the data?</td>
</tr>
<tr>
<td>• Who is responsible for data security?</td>
</tr>
<tr>
<td>• What, if any, sensitive patient data are being collected or used?</td>
</tr>
<tr>
<td>• What physical, technical, and procedural measures have been taken to safeguard patient data?</td>
</tr>
<tr>
<td>• What data safety and monitoring measures will be put in place?</td>
</tr>
<tr>
<td>• Do you and your partner have the relevant experience to serve as data stewards?</td>
</tr>
</tbody>
</table>

Conceptual Framework

Figure 3 shows the relationship among these 5 domains and how they might be leveraged to provide a risk stratification of global health data partnerships. This type of conceptual framework has been leveraged to address various issues in informatics, such as the development of global health networks, patient safety, and conceptual models for research [89-91]. The value of these conceptual frameworks is in laying out all the components of a given issue, exploring their interrelationships, and identifying the emerging complexity [92]. In the current framework, Demographics and Equity and Societal Freedom provide a foundational understanding of a given country. Economics of Health Care can be understood within that context, and the State of HIT is influenced by, and builds upon, all 3 domains. Partnership and Trust is a moderating factor for all other variables. A long-standing, effective partnership with high levels of trust and cooperation may overcome a number of deficiencies in other domains, whereas an unstable or ineffective partnership may suffer from serious data concerns despite an otherwise favorable environment. The latter case is often the cause for HIT implementation failures in high-income countries [93].

Figure 3. Four of the domains build on each other (Demographics and Equity, Economics of Health Care, Societal Freedom and Safety, and State of Health Information Technology). These are moderated by the fifth domain, Partnership and Trust. Together, these domains can be evaluated to produce a risk stratification for global health data partnerships and health information technology projects.

Discussion

Overview

As HIT deployments continue to progress in LMICs, data security concerns will become more prevalent. The development of this conceptual framework is an attempt to better understand the many variables that might affect health data security in a given country. There are a number of existing models for assessing the maturity of HIT and data security [94]. However, most of these models have been developed or applied in high-income countries and make assumptions about the legal, regulatory, and technical capacity already in place; these assumptions often do not hold true in LMICs. The health systems of high-income countries (and the social, political, and economic forces that support them) can vary significantly from
those in LMICs. For example, in the review by Tarhan et al [94], the authors provide the full list of countries in which the maturity models they reviewed were developed and applied. The average WHO health performance index of these countries was 0.813 (SD 0.17), whereas the average health performance index of CIS countries is 0.612 (SD 0.12; P=.001 by 2-tailed, 2 sample t test) [38]. Therefore, it may not be meaningful to apply existing maturity models to LMICs; to our knowledge, there are no models explicitly developed for low-resource settings.

The goal of our research is ultimately to develop an assessment and decision support tool that organizations can use in their global health partnerships. In the first version of this tool, we anticipate a more qualitative approach in which organizations use these domains to guide them in conducting a thorough evaluation of projects and partnerships. A second version of the tool will have a more quantitative component; 4 of the 5 key domains we identified use publicly available indices that could be integrated into a mathematical model that describes health data risk. Multimedia Appendix 5 presents the values for a sampling of indices across the domains of State of HIT, Economics of Health Care, Demographics and Equity, and Societal Freedom. We have included 9 CIS countries for which data are available (Armenia, Azerbaijan, Belarus, Kazakhstan, Kyrgyzstan, Moldova, Russia, Tajikistan, and Ukraine). In addition, data have been aggregated for several other countries to serve as reference points for the 4 domains of interest. These range from countries that have a track record of excellence in public health (Germany, France, Switzerland, South Korea, and Japan) to countries that have made significant advances in HIT (Estonia) and modernized their health care systems in other geographic regions (China and South Africa), as well as the United States. For the fifth domain, Partnership and Trust, additional research and validation will be required to finalize the self-assessment questionnaire.

The concept of bias was a recurring theme in many of the sources we evaluated. During the COVID-19 epidemic, the disproportionate impact that bias, discrimination, and racism can have on health outcomes was highlighted at scale [95]. These social biases can easily be translated into HIT; data sets, algorithms, and predictive models are all subject to both explicit and implicit bias [70]. The types of individuals who are included in data sets, the data elements that are and are not collected, the variables that are highlighted, and the outcomes that are selected for can all result in HIT applications that adversely affect care delivery and further drive health disparities among marginalized groups [70,96,97]. Although not directly linked to data security considerations (except to the extent to which security failures may adversely affect members of marginalized groups, as discussed previously in other domains), we believe that an awareness of bias is critical in any global health data partnership. Additional research is needed to identify approaches to measure and account for biases that may differ across settings.

Limitations

Our study includes several limitations. Although we reviewed the literature to inform our research and approach, there is still a need for a comprehensive systematic literature review to be conducted and published. Given the evolving nature of the subject, a scoping review methodology would be appropriate; our research team is preparing to embark on this project [98]. Our proposed domains reflect our research and experience but need further validation from the broader community of health care, HIT, and public policy professionals. Our conceptual framework has not been tested quantitatively or qualitatively against real-world examples; thus, it remains to be seen if it can meaningfully capture the complexities and nuances of health data security. Finally, it is unclear if our proposed indices will result in a useful quantitative model of risk; further analysis is required.

Next Steps

To advance our research agenda, we plan to engage in the following activities:

1. Validate our conceptual framework: Additional work is needed to validate our proposed framework. We will share our framework publicly to gather both formal and informal feedback from stakeholders around the world. We will also begin the work of applying the framework to real-world examples in collaboration with local experts to test its internal and external validity. Further literature review and qualitative research will be needed to finalize the Partnership and Trust self-assessment.

2. Develop a qualitative assessment tool: Once we have finalized and validated the framework, we will develop a qualitative assessment tool that other organizations can use to evaluate their existing data partnerships. This stage will not only provide additional external validation and refinement of the framework, but will also provide the opportunity to conduct user-centered design research to improve the usability of the tool and related documentation.

3. Develop a quantitative model: We plan to work with our data science colleagues to use both traditional statistical methods and more modern machine learning approaches to develop a quantitative model of our conceptual framework. This will require extensive validation, but if successful, it may result in a risk stratification that could conceivably be calculated for every country, needing only the Partnership and Trust self-assessment to provide local context.

4. Dissemination of findings: Our overall goal is to support how organizations make decisions around global health data partnerships. The current framework, assessment tools, and quantitative models are all intended to help organizations make the best decisions possible in terms of safeguarding patient data in LMICs. To that end, we intend to use multiple avenues to disseminate our work, including publications, presentations, webinars, and white papers. We plan to collaborate with universities and nongovernmental organizations to help them to implement the conceptual framework and associated tools.

Conclusions

Global health HIT partnerships have the potential to have a positive impact in LMICs, leveraging the resources and knowledge of partner organizations to build in-country capacity and expertise. However, there are gaps in the legal, technical,
and regulatory environments in many LMICs, increasing the risk of possible health data misuse, particularly among marginalized and vulnerable populations. Our conceptual framework helps to identify key domains that may have an impact on health data security considerations in global health partnerships. Additional research is needed to further validate and improve the framework. We encourage global health, HIT, and health care professionals to participate in improving this framework. In the future, we hope to be able to leverage our framework to create assessment and decision-making tools that can be used to evaluate risk in other global health initiatives such as clinical and academic partnerships, pandemic control, and emergency response operations.

Acknowledgments
The authors would like to acknowledge Mr Pushkar Sharma for his insights into quantitative indices of freedom and democracy around the world, Dr Laura Ferguson for her input on quality-of-care standards, and Dr Neeraj Sood and Mr Virat Agrawal for their help in refining the overall concepts presented and helping to identify additional quantitative indices. The authors would also like to express their gratitude to all the members of the Online Pediatric Education Network and Avetis team at Children’s Hospital Los Angeles for their camaraderie and support.

Authors’ Contributions
JE conceived of the overall study, outlined the manuscript, contributed to the writing and editing of the manuscript, and oversaw the project. ATS generated the first draft, performed the initial literature review, and assembled the textboxes. JD and TL contributed to the editing of the manuscript and provided feedback and subject matter expertise.

Conflicts of Interest
JE is a subject matter consultant for AI Health, a machine learning company that focuses on chronic disease management. JD holds equity in a health data company, Orderly Health. JD also serves as an advisor for an electronic medical record start up, Frontida Health.

Multimedia Appendix 1
State of health information technology domain.
[PDF File (Adobe PDF File), 17 KB - formative_v5i12e25833_app1.pdf]

Multimedia Appendix 2
Economics of healthcare domain.
[PDF File (Adobe PDF File), 34 KB - formative_v5i12e25833_app2.pdf]

Multimedia Appendix 3
Demographics and equity domain.
[PDF File (Adobe PDF File), 30 KB - formative_v5i12e25833_app3.pdf]

Multimedia Appendix 4
Societal freedom and safety domain.
[PDF File (Adobe PDF File), 29 KB - formative_v5i12e25833_app4.pdf]

Multimedia Appendix 5
Example index values for CIS and reference countries.
[PDF File (Adobe PDF File), 20 KB - formative_v5i12e25833_app5.pdf]

References


75. Bollyky T, Templin T, Cohen M, Schoder D, Dieleman J, Wigley S. The relationships between democratic experience, health outcomes, and political science. URL: https://personal.lse.ac.uk/martinak/ehealth.pdf [accessed 2021-01-21]


Abbreviations

CIS: Commonwealth of Independent States
EHR: electronic health record
GBD: Global Burden of Disease
GDP: gross domestic product
HIT: health information technology
LMIC: low- and middle-income country
WHO: World Health Organization

Edited by C Lovis; submitted 29.01.21; peer-reviewed by K Fuji, M Salman; comments to author 23.05.21; revised version received 30.06.21; accepted 10.10.21; published 08.12.21.

Please cite as:
Espinoza J, Sikder AT, Dickhomer J, Lee T
JMIR Form Res 2021;5(12):e25833
URL: https://formative.jmir.org/2021/12/e25833
doi: 10.2196/25833
PMID:
Test-Retest Reliability of Home-Based Fitness Assessments Using a Mobile App (R Plus Health) in Healthy Adults: Prospective Quantitative Study

I-I Lin¹, MSc; You-Lin Chen¹, MSc; Li-Ling Chuang²,³,⁴, PhD

¹Recovery Plus Inc, Chengdu, China
²School of Physical Therapy & Graduate Institute of Rehabilitation Science, College of Medicine, Chang Gung University, Taoyuan, Taiwan
³Department of Physical Medicine and Rehabilitation, Chang Gung Memorial Hospital at Linkou, Taoyuan, Taiwan
⁴Healthy Aging Research Center, Chang Gung University, Taoyuan, Taiwan

Corresponding Author:
Li-Ling Chuang, PhD
School of Physical Therapy & Graduate Institute of Rehabilitation Science, College of Medicine, Chang Gung University
No. 259 Wen-hua 1st Rd, Guishan Dist
Taoyuan, 33302
Taiwan
Phone: 886 3 2118800 ext 3177
Fax: 886 3 2118700
Email: lchuang@gap.cgu.edu.tw

Abstract

Background: Poor physical fitness has a negative impact on overall health status. An increasing number of health-related mobile apps have emerged to reduce the burden of medical care and the inconvenience of long-distance travel. However, few studies have been conducted on home-based fitness tests using apps. Insufficient monitoring of physiological signals during fitness assessments have been noted. Therefore, we developed R Plus Health, a digital health app that incorporates all the components of a fitness assessment with concomitant physiological signal monitoring.

Objective: The aim of this study is to investigate the test-retest reliability of home-based fitness assessments using the R Plus Health app in healthy adults.

Methods: A total of 31 healthy young adults self-executed 2 fitness assessments using the R Plus Health app, with a 2- to 3-day interval between assessments. The fitness assessments included cardiorespiratory endurance, strength, flexibility, mobility, and balance tests. The intraclass correlation coefficient was computed as a measure of the relative reliability of the fitness assessments and determined their consistency. The SE of measurement, smallest real difference at a 90% CI, and Bland–Altman analyses were used to assess agreement, sensitivity to real change, and systematic bias detection, respectively.

Results: The relative reliability of the fitness assessments using R Plus Health was moderate to good (intraclass correlation coefficient 0.8-0.99 for raw scores, 0.69-0.99 for converted scores). The SE of measurement and smallest real difference at a 90% CI were 1.44-6.91 and 3.36-16.11, respectively, in all fitness assessments. The 95% CI of the mean difference indicated no significant systematic error between the assessments for the strength and balance tests. The Bland–Altman analyses revealed no significant systematic bias between the assessments for all tests, with a few outliers. The Bland–Altman plots illustrated narrow limits of agreement for upper extremity strength, abdominal strength, and right leg stance tests, indicating good agreement between the 2 assessments.

Conclusions: Home-based fitness assessments using the R Plus Health app were reliable and feasible in young, healthy adults. The results of the fitness assessments can offer a comprehensive understanding of general health status and help prescribe safe and suitable exercise training regimens. In future work, the app will be tested in different populations (eg, patients with chronic diseases or users with poor fitness), and the results will be compared with clinical test results.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2000030905; http://www.chictr.org.cn/showproj.aspx?proj=50229

(JMIR Form Res 2021;5(12):e28040) doi:10.2196/28040
**KEYWORDS**
mobile health app; reliability; home-based fitness assessments; healthy adults; mobile phone; digital health

**Introduction**

**Background**

Physical fitness plays an important role in overall health and quality of life and is directly related to physical activity [1]. Regular physical activity confers health benefits, such as increased life expectancy and reduced mortality [2,3]. However, the World Health Organization has reported that >80% of adolescents globally do not engage in sufficient physical activity. The prevalence of insufficient physical activity was 27.5% among adults aged >18 years worldwide [4]. Studies have indicated that physical inactivity is associated with poor physical fitness and increases not only the incidence and mortality rates of chronic disease, but also the medical and economic burden of disease [5-7]. Physical fitness has various degrees of influence on the activities and quality of life [1,8]. Poor physical fitness (below the 25th percentile of the fitness distribution) has a much greater impact on the risk of cardiovascular disease than insufficient physical activity [9]. Therefore, physical fitness needs to be considered as a fundamental assessment for people with a higher risk of chronic diseases.

Several physical fitness assessment methods have been established for reliability and validity. Physical fitness measures typically consist of cardiorespiratory fitness, muscle strength, endurance, agility, flexibility, and measures of body composition [1,10,11]. The 3-minute step test is one of the common cardiorespiratory fitness tests, consisting of stepping up and down a height of 23.0 cm-50.8 cm at a consistent step rate [12]. The 3-minute step test was shown to be reliable and valid in the general population and in patients with lung disease and rheumatoid arthritis [12-17]. Sufficient muscle power and endurance can reduce the risk of exercise injury and enhance cardiorespiratory capacity [18-20]. Wall squatting, push-up, and curl-up tests are common strength tests for the lower and upper extremities and abdominal muscles with established validity and reliability [11,21-23]. Balance and flexibility are important because poor stability may increase the risk of falls and limit functional activities [24-27]. Insufficient flexibility and mobility may restrict movement and cause pain [28-30]. Balance tests, the toe-touch test, the sit-and-reach test, and the Apley shoulder scratch test are common tests to assess balance, flexibility, and mobility [26,31-33]. However, most of these fitness tests are administered by a professional face to face, so patients or clients need to be present at a clinic or gym.

In consideration of cost and travel barriers, self-administered and home-based fitness tests may be more suitable for many people. However, there is currently little research on home-based fitness tests. One study showed that the home-based Senior Fitness Test, using inertia sensors and a depth camera, led to greater leg or arm strength, aerobic endurance, and flexibility [29]. The InterWalk Fitness Test incorporates indirect calorimetry and acceleration monitoring and was found to be accurate and reliable for persons with type 2 diabetes [34]. The self-administered Canadian Home Fitness Test was developed to assess cardiorespiratory endurance with a double 8-inch step and has an established record of safety and predictive ability [35-37]. Additional reliable home-based fitness tests that are easy to use and record data on accessible software platforms are needed.

As mobile technologies have advanced, an increasing number of health-related apps have emerged [38]. Some health apps provide patient education about lifestyle and health behaviors, some provide pain management, and others provide physical fitness assessments or interventions [34,38-40]. Among commercial fitness apps, most focus on cardiorespiratory fitness assessments, such as the submaximal walking data collected by a smartphone’s accelerometer [40]. Some apps focus on functional performance, such as movement speed or leg strength during functional activities [41]. However, most commercial fitness apps lack supporting evidence [40]. Only a few fitness apps have been tested for validity and reliability, and most are rated as having moderate to good validity [34,42-44]. Insufficient monitoring of physiological signals (eg, heart rate) during cardiorespiratory fitness assessment was noted among the available apps [40]. Therefore, we designed R Plus Health (Recovery Plus Inc), a digital health app that incorporates all the usual components of a fitness assessment but also monitors physiological signals.

**Objective**

The aim of this study is to investigate the test-retest reliability of home-based fitness assessments using the R Plus Health app in healthy young adults.

**Methods**

**Participants**

A total of 31 participants were recruited with convenience sampling from 4 departments of a technology company in Chengdu, China. Sampling was performed via random draw. The inclusion criteria were healthy adults with normal health examinations, aged between 18 and 75 years, and with the ability to use smartphones. Those who rated more than 3 out of 10 on the visual analog pain scale; had poor compliance or were not willing to cooperate with the assessment; had regular strengthening sessions during the study period; had a history of alcohol abuse or illegal drug use; were pregnant, lactating, or trying to become pregnant; had participated in other clinical trials within 3 months before this study; and had uncontrolled chronic diseases were excluded. The participants received oral and written information about the study, and informed consent was obtained from all participants. This study was approved by the Chinese Ethics Committee of Registering Clinical Trials (ChiCTR2000030905).

**R Plus Health App**

The R Plus Health app was developed as a tool for healthy adults and patients with chronic diseases. It includes fitness assessments and individualized exercise prescriptions with physiological signal monitors (eg, heart rate monitor). After...
downloading the R Plus Health app, the participants received an informed safety declaration and completed a health questionnaire, which was checked by doctors or other professional health care providers on the web. Through oral and video guidance, the participants were then instructed on how to perform the fitness assessments with maximal effort. The fitness assessments in the app included cardiorespiratory fitness, strength, balance, mobility, and flexibility tests (Figure 1). These fitness assessments have established clinical validity and reliability [12-25,30-33]. To complete the cardiorespiratory fitness test and record a real-time heart rate, the participants were required to wear a heart rate monitor below the sternum on a strap around the chest during testing (Figure 2). The heart rate monitor (Magene H64 dual protocol heart rate sensor) is compatible with the app and has Conformite Europeenne and Federal Communications Commission certification. Finally, according to the results of the fitness assessments and the overall health condition of each participant, a proper individualized exercise prescription was suggested by professional teams in the app.

**Figure 1.** Video demonstration of the push-up test.

![Push-up Test](image1)

**Figure 2.** Demonstration of how to wear the heart rate monitor strap.

![Heart Rate Monitor](image2)
Assessment Procedures

Eligible participants enrolled in the study, provided informed consent, downloaded the R Plus Health app, and filled in the health questionnaire with the assistance of a physiotherapist. The physiotherapist recorded the basic data, including pain level on a visual analog scale and the overall health condition of the participants, at the baseline and final assessments.

All participants self-administered 2 fitness assessments with a 2- to 3-day interval between assessments to provide the best reproducibility [45]. The fitness assessments were administered sequentially (cardiorespiratory endurance, strength, flexibility, mobility, and balance). The participants followed the guidance and instructions in the app for each fitness assessment.

The 3-minute step test measures cardiorespiratory endurance based on how quickly the heart rate returns to normal after a 3-minute step exercise [12,13]. First, the heart rate monitor strap was worn for a 5-minute rest period beside the 30-cm step (to establish a baseline). After watching the tutorial videos in the app, the participants stepped up and down at 96 beats per minute (bpm) using a metronome for a total of 3 minutes. After finishing the test, the participants rested for 1 minute. The participants could suspend the test if any discomfort occurred.

The push-up and curl-up tests measure muscle strength and endurance in the upper limbs and abdomen, respectively, based on the number of completed repetitions [11,21,23,46]. When performing the push-up test, there were 2 variations in the starting position. The standard push-up test involved having the knees off the ground in the push-up position and was used for male participants. The modified push-up test involved having the knees on the ground and was used for female participants. The participants performed as many push-ups as possible with the correct form within 40 seconds. The curl-up test began with the participants lying on their back, knees bent at approximately 90°, feet flat on the floor, and arms straight with the palms of their hands resting on their thighs. The participants curled up and down at 40 bpm using a metronome. If the participants could not continue or stopped for more than 5 seconds, they clicked the completed button and recorded the repetitions.

The wall squatting test measures muscle strength and endurance in the lower limbs based on the holding time [11,22]. The wall squatting test began with the participants in a standing position, feet shoulder width apart and back against the wall; then, both knees were bent at a 90° angle. The participants held this squatting position for as long as possible. When the participants were finished, they could click the completed button and record the total time. If the participants held the position for more than 150 seconds, the app finished the test automatically.

The sit-and-reach test measures the flexibility of the hamstrings and the lower back with a ruler based on the distance [11,30,32]. The participants sat on the floor with their legs straight and their heels in line with a ruler, hands stacked, and palms facing downward. They then reached forward as far as possible along the measuring line. After reaching forward, the participants recorded their distance in centimeters.

The Apley scratch test or the upper extremity (UE) multipattern test measures the mobility of the upper limbs based on the distance between the middle fingers [33]. There were 2 patterns of upper limb flexibility: shoulder flexion, abduction, and external rotation and shoulder extension, adduction, and internal rotation. The participants performed these 2 patterns of movement for each upper limb and recorded the distance between both middle fingers. The results were classified as above average, normal, or below average.

The one-leg stance test measures balance based on the holding time [26]. The participants stood on one leg, bent the other leg 15-20 cm off the ground with their eyes open and their arms beside the hips. The participants maintained their balance for as long as possible. If the participants lost balance, they clicked the completed button, and the time was recorded in the app automatically. If the participants maintained balance for more than 30 seconds, the app finished the test automatically.

To minimize possible diurnal variation in physical fitness, the participants were instructed to perform the 2 assessments at the same time of the day. They were asked to avoid resistance training and exhausting work between assessments to minimize the potential effects of fatigue. After each test, the participants immediately recorded the results on paper to avoid recall effects and then sent them to the researchers. The researchers concealed the data of the participants in an envelope for anonymity and encoded the names as numbers to protect the privacy of the participants.

Outcome Measures

At the baseline assessment, the descriptive data, pain score, and health condition of the participants were evaluated by a physiotherapist. Descriptive data included age, sex, height, and weight. The pain level was assessed using a visual analog scale from 0 (no pain) to 10 (worst pain). Health condition was assessed using a health-related questionnaire in the app and by a physiotherapist.

The outcomes of each fitness assessment included the raw data and converted score. The raw data were recorded as bpm, repetitions, seconds, and an ordinal scale. The converted scores (0-100) were computed using the app through normative data and a self-established score conversion system on the basis of the raw data.

The participants recorded the heart rate in bpm as raw data, and the converted scores used the same units. If someone could not complete the 3-minute step test, the reason was noted [12]. In each cardiorespiratory fitness test, 2 measurements were made: the average resting heart rate during the 5-minute rest period and the 1-minute recovery heart rate after the 3-minute step test.

The outcomes of the push-up, wall squatting, and curl-up tests were recorded as completed repetitions and total time. The flexibility of the lower limbs and lower back was measured in centimeters from negative to positive values. The mobility of the upper limbs was classified as above average, normal, or below average. The one-leg stance test recorded the total time in seconds [26,31-33].

Data Analysis

Statistical analyses were conducted using SPSS 20.0 software (IBM Inc). Descriptive statistics were presented in the form of
mean and SD, and the relative and absolute test-retest reliabilities of the fitness assessments were estimated separately.

**Relative Reliability**
The relative reliability of the fitness assessments was calculated using the intraclass correlation coefficient (ICC) with a 2-way mixed model (type absolute agreement). On the basis of the 95% CIs of the ICC estimates, agreement was rated as poor (<0.5), moderate (between 0.5 and 0.75), good (between 0.75 and 0.9), or excellent (>0.9) [47].

**Absolute Reliability**
The absolute reliability of the fitness assessments was evaluated using the SE of measurement (SEM), the smallest real difference (SRD), and Bland–Altman analyses [47,48]. The SEM expressed the measurement error variation between the assessments within a group and was calculated as \( SD_{pooled} \times \sqrt{(1-ICC)} \) [49]. In this formula, \( SD_{pooled} \) indicates the pooled SD for the 2 assessments. The SRD is a measure of sensitivity to change, represented as the magnitude of the change detected at a certain CI [50]. The \( SRD_{90} \) is defined as the SEM of the difference scores at a 90% confidence level and was calculated as \( 1.65\times\sqrt{2}\times SEM \) [48]. If the difference between the 2 assessments was greater than the SRD, it was interpreted as a real change. For all measurements, the smaller the SEM and \( SRD_{90} \), the greater the reliability.

**Results**

**Overview**
The characteristics of the participants and the descriptive statistics of the fitness assessments at baseline are shown in Tables 1 and 2, respectively. The study enrolled 31 participants (Table 1), which exceeded the minimum sample size of 26 (effect size of 0.5 and power of 0.8) calculated using G*power 3.1 [54]. The average age was 27.25 (4.0) years, and they had negligible pain (mean 0.19 out of 10 on the visual analog scale), which did not worsen during testing.

**Table 1. Characteristics of the participants (N=31).**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>27.25 (4.0)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>16 (52)</td>
</tr>
<tr>
<td>Male</td>
<td>15 (48)</td>
</tr>
<tr>
<td>Height (cm), mean (SD)</td>
<td>168.66 (7.61)</td>
</tr>
<tr>
<td>Weight (kg), mean (SD)</td>
<td>60.23 (11.41)</td>
</tr>
<tr>
<td>BMI (kg/m(^2)), mean (SD)</td>
<td>21.03 (2.75)</td>
</tr>
<tr>
<td>Health status</td>
<td>Normal health examination</td>
</tr>
<tr>
<td>Pain assessment (range 0-10), mean (SD)</td>
<td>0.19 (0.65)</td>
</tr>
</tbody>
</table>
Table 2. Fitness assessments of the participants at baseline (N=31).

<table>
<thead>
<tr>
<th>Domain and test items</th>
<th>Raw data, mean (SD)</th>
<th>Converted score&lt;sup&gt;a&lt;/sup&gt;, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cardiovascular fitness</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HR&lt;sup&gt;b&lt;/sup&gt; at rest&lt;sup&gt;c&lt;/sup&gt; (bpm&lt;sup&gt;d&lt;/sup&gt;)</td>
<td>74.81 (9.6)</td>
<td>51.23 (16.5)</td>
</tr>
<tr>
<td>1-minute HR after test&lt;sup&gt;e&lt;/sup&gt; (bpm)</td>
<td>92.26 (18.3)</td>
<td>60.19 (19.1)</td>
</tr>
<tr>
<td>UE&lt;sup&gt;f&lt;/sup&gt; strength: push-up (repetitions)</td>
<td>12.94 (9.3)</td>
<td>58.71 (18.3)</td>
</tr>
<tr>
<td>Abdominal strength: curl-up (repetitions)</td>
<td>19.55 (13.7)</td>
<td>51.29 (18.8)</td>
</tr>
<tr>
<td>LE&lt;sup&gt;g&lt;/sup&gt; strength: wall squatting (seconds)</td>
<td>63.03 (26.1)</td>
<td>53.23 (18.1)</td>
</tr>
<tr>
<td>LE flexibility: sit-and-reach (centimeters)</td>
<td>2.85 (14.2)</td>
<td>57.74 (25.3)</td>
</tr>
<tr>
<td><strong>Balance ability</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right leg stance (seconds)</td>
<td>31.77 (14.4)</td>
<td>—&lt;sup&gt;h&lt;/sup&gt;</td>
</tr>
<tr>
<td>Left leg stance (seconds)</td>
<td>30.55 (9.9)</td>
<td>—</td>
</tr>
<tr>
<td><strong>UE mobility&lt;sup&gt;i&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UE multipattern (above average), n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right UE</td>
<td>23 (74)</td>
<td></td>
</tr>
<tr>
<td>Left UE</td>
<td>16 (52)</td>
<td></td>
</tr>
<tr>
<td>UE multipattern (normal), n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right UE</td>
<td>3 (10)</td>
<td></td>
</tr>
<tr>
<td>Left UE</td>
<td>6 (19)</td>
<td></td>
</tr>
<tr>
<td>UE multipattern (below average), n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right UE</td>
<td>5 (16)</td>
<td></td>
</tr>
<tr>
<td>Left UE</td>
<td>9 (29)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Converted score (0-100) from raw data in the app using normative data.

<sup>b</sup>HR: heart rate.

<sup>c</sup>Resting heart rate measurement.

<sup>d</sup>bpm: beats per minute.

<sup>e</sup>Heart rate recovery 1 minute after the 3-minute step test.

<sup>f</sup>UE: upper extremity.

<sup>g</sup>LE: lower extremity.

<sup>h</sup>Not available; no converted score was calculated respectively because the scores were averaged in balance ability.

<sup>i</sup>The upper extremity multipattern test was categorized into 3 classes (above average, normal, and below average).

Table 2 shows the results of the baseline fitness assessments as raw data (mean [SD]) and converted score (0-100). At the baseline assessments, the average heart rate was 74.81 bpm at rest, and the 1-minute recovery heart rate was 92.26 bpm after the 3-minute step test. In the strength tests, the average number of completed repetitions was 12.94 push-ups and 19.55 curl-ups, and the average holding time for the squatting test was 63.03 seconds.

**Relative Reliability**

Table 3 summarizes the test-retest reliability of all the fitness assessments. On the basis of the raw data, the ICCs for all tests were 0.8-0.99. On the basis of the converted scores, the ICCs for all tests were 0.69 to 0.99. In most tests, the 95% CI was >0.5.
Table 3. Test-retest reliability of the fitness assessments (N=31).

<table>
<thead>
<tr>
<th>Test items</th>
<th>ICC$^a$ for the raw data</th>
<th>ICC for the converted score</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR$^b$ at rest$^c$</td>
<td>0.80 (0.58-0.90)</td>
<td>0.69 (0.34-0.85)</td>
</tr>
<tr>
<td>1-minute HR after test$^d$</td>
<td>0.92 (0.84-0.96)</td>
<td>0.82 (0.63-0.92)</td>
</tr>
<tr>
<td>UE strength$^e$</td>
<td>0.97 (0.94-0.99)</td>
<td>0.97 (0.93-0.99)</td>
</tr>
<tr>
<td>Abdominal strength$^f$</td>
<td>0.98 (0.95-0.99)</td>
<td>0.94 (0.87-0.97)</td>
</tr>
<tr>
<td>LE strength$^g$</td>
<td>0.93 (0.85-0.96)</td>
<td>0.82 (0.63-0.92)</td>
</tr>
<tr>
<td>LE flexibility$^h$</td>
<td>0.89 (0.77-0.95)</td>
<td>1</td>
</tr>
<tr>
<td>UE mobility$^i$</td>
<td>N/A$^j$</td>
<td>0.99 (0.98-0.99)</td>
</tr>
<tr>
<td>Right leg stance</td>
<td>0.99 (0.98-0.99)</td>
<td>0.75 (0.5-0.88)</td>
</tr>
<tr>
<td>Left leg stance</td>
<td>0.89 (0.77-0.95)</td>
<td>0.75 (0.5-0.88)</td>
</tr>
</tbody>
</table>

$^a$ ICC: intraclass correlation coefficient (at a 95% CI).
$^b$ HR: heart rate.
$^c$ Resting heart rate measurement.
$^d$ 1-minute HR after test: heart rate recovery 1 minute after the 3-minute step test.
$^e$ UE strength: upper extremity strength (push-up test).
$^f$ Curl-up test.
$^g$ LE strength: lower extremity strength (wall squatting test).
$^i$ UE mobility: upper extremity mobility (upper extremity multipattern test).
$^j$ N/A: not applicable; no intraclass correlation coefficient value was calculated because the raw data of the upper extremity mobility test was the percentage of participants, not a continuous variable.

**Absolute Reliability**

The absolute reliability and Bland–Altman analyses are presented in Table 4. The SEM and SRD$^{90}$ were 1.44-6.91 and 3.36-16.11, respectively, across the different assessments. The mean differences in UE strength, lower extremity (LE) flexibility, and right leg balance tests were close to 0. The 95% CI of the mean difference contained 0, indicating no significant systematic error between the 2 assessments in strength (−6.28 to 3.89 in the LE strength test and −1.54 to 0.89 in the UE strength test), flexibility (−2.65 to 3.57 in the LE flexibility test), and balance tests (−1.75 to 0.07 in the right leg stance test and −5.58 to 0.93 in the left leg stance test).
Table 4. Absolute reliability of the fitness assessments in raw data.

<table>
<thead>
<tr>
<th>Raw data of test items</th>
<th>SEM&lt;sup&gt;a&lt;/sup&gt;</th>
<th>SRD&lt;sub&gt;90&lt;/sub&gt;&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Bland–Altman analyses</th>
<th>95% CI of d&lt;sup&gt;e&lt;/sup&gt;</th>
<th>LOA&lt;sup&gt;f&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR at rest&lt;sup&gt;g&lt;/sup&gt; (bpm&lt;sup&gt;h&lt;/sup&gt;)</td>
<td>4.28</td>
<td>9.99</td>
<td>5.61</td>
<td>5.58</td>
<td>1.00</td>
</tr>
<tr>
<td>1-minute HR after test&lt;sup&gt;i&lt;/sup&gt; (bpm)</td>
<td>5.18</td>
<td>12.08</td>
<td>7.19</td>
<td>6.91</td>
<td>1.24</td>
</tr>
<tr>
<td>UE strength&lt;sup&gt;j&lt;/sup&gt; (repetitions)</td>
<td>1.61</td>
<td>3.76</td>
<td>−0.32</td>
<td>3.31</td>
<td>0.59</td>
</tr>
<tr>
<td>Abdominal strength&lt;sup&gt;k&lt;/sup&gt; (repetitions)</td>
<td>2.08</td>
<td>4.85</td>
<td>−1.74</td>
<td>4.00</td>
<td>0.72</td>
</tr>
<tr>
<td>LE strength&lt;sup&lt;l&gt;&lt;/sup&gt; (s)</td>
<td>6.91</td>
<td>16.11</td>
<td>−1.19</td>
<td>13.87</td>
<td>2.49</td>
</tr>
<tr>
<td>LE flexibility&lt;sup&gt;m&lt;/sup&gt; (cm)</td>
<td>4.71</td>
<td>10.99</td>
<td>0.46</td>
<td>8.47</td>
<td>1.52</td>
</tr>
<tr>
<td>Right leg stance (s)</td>
<td>1.44</td>
<td>3.36</td>
<td>−0.84</td>
<td>2.48</td>
<td>0.45</td>
</tr>
<tr>
<td>Left leg stance (s)</td>
<td>3.28</td>
<td>7.66</td>
<td>−2.32</td>
<td>8.87</td>
<td>1.59</td>
</tr>
</tbody>
</table>

<sup>a</sup>SEM: SE of measurement.

<sup>b</sup>SRD<sub>90</sub>: smallest real difference at a 90% confidence level.

<sup>c</sup>d: mean difference between 2 trials.

<sup>d</sup>SD<sub>diff</sub>: SD of mean difference.

<sup>e</sup>SD<sub>diff</sub>/√n.

<sup>f</sup>LOA: limits of agreement (d±[SD<sub>diff</sub>×1.96]).

<sup>g</sup>HR at rest: resting heart rate measurement.

<sup>h</sup>bpm: beats per minute.

<sup>i</sup>1-minute HR after test: heart rate recovery in 1 minute after the 3-minute step test.

<sup>j</sup>UE strength: upper extremity strength (push-up test).

<sup>k</sup>Abdominal strength: curl-up test.

<sup>l</sup>LE strength: lower extremity strength (wall squatting test).

<sup>m</sup>LE flexibility: lower extremity flexibility (sit-and-reach test).

Figures 3-10 show the Bland–Altman plots of the differences between the 2 measurements for all tests. Reference lines show mean differences between time 1 and time 2 (solid line) and 95% LOA for the mean difference (dotted lines). The differences for most of the tests were within the 95% CI. The LOA were −5.32 to 16.55 for the heart rate at rest and −6.34 to 20.73 for the 1-minute heart rate after test. The LOA were −6.81 to 6.17 in the UE strength test, −9.58 to 6.10 in the abdominal strength test, and −28.38 to 25.99 in the LE strength test. The LOA were −16.14 to 17.06 in the LE flexibility test, −5.70 to 4.02 in the right leg stance test, and −19.70 to 15.06 in the left leg stance test. There were at most 3 outliers in the 1-minute heart rate after, LE strength, and right leg stance tests.
**Figure 3.** The Bland–Altman plots of differences between the 2 measurements in heart rate at rest. HR: heart rate.

**Figure 4.** The Bland–Altman plots of differences between the 2 measurements in 1-minute heart rate recovery. HR: heart rate.
Figure 5. The Bland–Altman plots of differences between the 2 measurements in abdominal strength assessments. ab: abdominal.

Figure 6. The Bland–Altman plots of differences between the 2 measurements in upper extremity strength assessments. UE: upper extremity.
Figure 7. The Bland–Altman plots of differences between the 2 measurements in lower extremity strength assessments. LE: lower extremity.

Figure 8. The Bland–Altman plots of differences between the 2 measurements in LE flexibility tests. LE: lower extremity.
Figure 9. The Bland–Altman plots of differences between the 2 measurements in right leg stance tests. R: right.

Figure 10. The Bland–Altman plots of differences between the 2 measurements in left leg stance tests. L: left.
Discussion

Principal Findings

This is the first study to investigate the test-retest reliability of home-based fitness assessments using a mobile health app in young, healthy adults. The results showed a moderate to good reliability of the fitness assessments. Therefore, through video and oral guidance, the app was shown to be reliable when applied to young users.

The self-administered fitness assessments in the app were feasible, with a low risk of injury. All participants completed the fitness assessments with the guidance of the R Plus Health app. Although some participants enrolled in the study had mild pain, they did not worsen after the fitness assessment. In other clinical research, it has been shown that mobile apps are able to conduct ecological momentary assessments, manage, and monitor patients with good adherence, detect symptoms, and evaluate the condition of a patient [55-57]. Therefore, well-designed mobile apps could offer a feasible means of self-assessment for clients and clinicians.

The results of our study were consistent with those of previous reliability studies [14]. Among the fitness assessments in the app, the test-retest reliabilities were moderate to good in this study. On the basis of the raw data, the ICCs for all tests were 0.8-0.99, indicating good to excellent reliability. On the basis of the converted scores, the ICCs for all tests were 0.68-0.99, indicating moderate to good reliability. The 95% CIs were above 0.5 in most tests. One previous study investigated the reliability of web-based versus supervised cardiovascular fitness assessments using the Young Men’s Christian Association 3-minute step test for college students [14]. The results of that study showed that there were no significant differences in the recovery heart rate between the 2 groups and that self-assessed cardiovascular fitness measurements were reliable [14]. Another study investigating the reliability of the Chester Step Test in patients with chronic obstructive pulmonary disease showed good reliability (ICC>0.8) [58]. In a previous analysis of strength fitness tests, reliability was established in adolescents, with ICCs of 0.7-0.9 in push-up, curl-up, and wall squatting tests [59]. For balance tests, the ICCs of single-leg stance tests were found to be >0.77 in young adults using a computerized balance platform [60]. These findings suggest that, regardless of the methods of assessing fitness capacity (eg, web-based and supervised assessments), the use of standard procedures and precise guidance under signal monitoring can ensure an accurate measure of actual performance. Self-administered fitness assessments in the R Plus Health app can be one of these efficient and reliable methods.

In addition to the relative reliability, the absolute reliability can demonstrate the agreement and sensitivity of the mean differences between the assessments. The SRD is a measure of sensitivity to change and represents the magnitude of the change at a certain confidence level [50]. If the difference between 2 assessments was larger than the SRD, it could be considered a real change, and the smaller the SEM and SRD90 of the difference, the greater the reliability. In this study, the SEM and SRD90 ranged from 1.44-6.91 and 3.36-16.11, respectively. For example, if the change was more than 16.11 in the wall squatting test, it was considered real at a 90% confidence level. In this study, the SRD90 in the wall squatting, push-up, and curl-up tests were 16.11, 3.76, and 4.85, respectively. These values were greater than the between-assessment changes reported in a previous study, which were 6.2, 2.6, and 0.1 for the wall squatting, push-up, and curl-up tests, respectively [59]. Different results might be because of different populations, ages, and experimental designs.

The Bland–Altman analyses and plots were generated to measure the repeatability of 2 measurement systems or of several trials using one method [49,51]. The scattering of data points within the 95% LOA and a smaller range between the 2 limits indicated higher agreement [52,53]. The 95% CI of the mean difference contained 0, indicating no significant systematic error between the 2 assessments for the strength, flexibility, and balance tests. The range of the LOA was slightly narrower in the UE strength (−6.81 to 6.17), abdominal muscle strength (−9.58 to 6.10), and right leg stance (−5.70 to 4.02) tests, indicating higher agreement. There was at least one outlier in each fitness assessment, and at most 3 outliers (in the 1-minute heart rate, LE strength, and right leg stance tests), which might be due to familiarization or fatigue in the second test.

Sufficient physical fitness is critical in daily life. It can decrease the risk of cardiovascular disease, pain, and injuries and improve the performance of life activities [9,18-20]. From the results of the cardiorespiratory fitness assessments in this study, the mean heart rate after 1-minute recovery from the step test was 92.26 bpm, indicating above average fitness base on the normative data [61]. In the LE strength wall squatting test, the mean holding time was 63.03 seconds, indicating an average fitness level [62]. In the LE flexibility sit-and-reach test, the mean distance was 2.85 cm, categorized as an average fitness level [11]. Even though the enrolled participants were generally in good health, they were below average in some of the fitness components. In the push-up test for UE strength and the curl-up test for abdominal muscle strength test, the mean number of repetitions was 12.94 and 19.55, respectively, which was below average based on the normative data and showed a need for improvement [63,64]. In the single-leg stance balance test, the mean holding time was 30 seconds (31.77 seconds for the right leg and 30.55 seconds for the left leg), indicating a below average fitness level [65]. Lack of muscle strength and balance can increase the risk of falls, pain, and injuries and limit daily life activities [24-27]. Therefore, comprehensive fitness assessments are essential.

Each participant differed in their performance in the physical fitness assessments according to variable self-conditions between the 2 assessments, and the results also differed from one participant to another. Even in healthy participants without chronic diseases, mild pain can lead to low strength in the extremities. Pain can inhibit muscle firing, and the lack of muscle contraction can decrease the stability of the joints and in turn, produce pain [66]. In other situations, insufficient muscle strength can lead to poor cardiorespiratory fitness. Evidence has shown that muscular fitness is related to cardiovascular prognosis and mortality [67]. As a result, according to the
individual situation, it is important to detect weaknesses in the fitness profile and provide proper assessments and advice to clients. Through a comprehensive fitness assessment composed of cardiovascular endurance, strength, flexibility, and balance tests, the R Plus Health app can provide clinicians with a complete picture of the clients’ fitness. Clinicians can then choose to provide other detailed assessments on the web or at the clinic, which not only increases the efficiency of the evaluations but also decreases the medical and economic burden.

Limitations and Future Studies

This study had several limitations. First, the level of difficulty in similar assessments differed from one participant to another, which may lead to ceiling or floor effects. According to the individual situation, automatic adjustment of the grade of assessments will be essential. Second, it is difficult to ascertain the accuracy of the performance assessments in the app without professional supervision. That is, the results of the app might not be identical to testing under professional supervision. In this study design, the results from different testing situations (with or without supervision) could not be compared. One solution to this problem would be to apply suitable monitoring equipment (eg, motion capture analysis and artificial intelligence techniques) to increase the precision of the assessments. However, this creates an additional technological burden. Cross-validation of the outcomes collected by the app versus professional staff will be the subject of future studies. Third, the study recruited young, healthy adults, so the results of the fitness assessments should not be generalized to other populations, such as older adults or patients with chronic diseases. Therefore, the fitness assessments in the app need to be conducted in other populations to compare the results between the app and clinical testing. Testing of the R Plus Health app in additional populations will be conducted in the future.

Conclusions

Home-based fitness assessments using a mobile health app were reliable and feasible in young, healthy adults. The results showed moderate to good reliability, and the testing process caused negligible pain effects. This study highlighted an important contribution of mobile health apps to health care, that is, that healthy adults can self-administer fitness tests and thereby reduce overall costs. The results of mobile fitness assessments can offer a reliable understanding of a person’s health condition and help prescribe a safe and suitable exercise training regimen. Expansion of the use of this technology to different populations (eg, patients with chronic diseases or users with poor fitness) will offer widespread benefits to both patients and the health care system.

Acknowledgments

This research was partially supported by the Ministry of Science and Technology (MOST-109-2314-B-182-030, and 110-2314-B-182-017) and Chang Gung Memorial Hospital (CMRPD110141 and CMRPD 110142) in Taiwan.

The authors thank Si-jing Ye and Jing Wei for data collection and Chong Jiang for assistance in conducting the survey. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References


Abbreviations

- bpm: beats per minute
- ICC: intraclass correlation coefficient
- LE: lower extremity
- LOA: limits of agreement
- SEM: SE of measurement
- UE: upper extremity

©I-I Lin, You-Lin Chen, Li-Ling Chuang. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.12.2021. 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.
Adoption and Appropriateness of mHealth for Weight Management in the Real World: A Qualitative Investigation of Patient Perspectives

Jessica Y Breland¹, PhD; Khizran Agha¹, MBBS; Rakshitha Mohankumar¹, BA
Center for Innovation to Implementation, VA Palo Alto Health Care System, Menlo Park, CA, United States

Corresponding Author:
Jessica Y Breland, PhD
Center for Innovation to Implementation
VA Palo Alto Health Care System
795 Willow Road (MPD-152)
Menlo Park, CA, 94025
United States
Phone: 1 650 493 5000
Email: jessica.breland@va.gov

Abstract

Background: Mobile health (mHealth) interventions for weight management can result in weight loss outcomes comparable to in-person treatments. However, there is little information on implementing these treatments in real-world settings.

Objective: This work aimed to answer two implementation research questions related to mHealth for weight management: (1) what are barriers and facilitators to mHealth adoption (initial use) and engagement (continued use)? and (2) what are patient beliefs about the appropriateness (ie, perceived fit, relevance, or compatibility) of mHealth for weight management?

Methods: We conducted semistructured interviews with patients with obesity at a single facility in an integrated health care system (the Veterans Health Administration). All participants had been referred to a new mHealth program, which included access to a live coach. We performed a rapid qualitative analysis of interviews to identify themes related to the adoption of, engagement with, and appropriateness of mHealth for weight management.

Results: We interviewed 24 veterans, seven of whom used the mHealth program. Almost all participants were ≥45 years of age and two-thirds were White. Rapid analysis identified three themes: (1) coaching both facilitates and prevents mHealth adoption and engagement by promoting accountability but leading to guilt among those not meeting goals; (2) preferences regarding the mode of treatment delivery, usability, and treatment content were barriers to mHealth appropriateness and adoption, including preferences for in-person care and a dislike of self-monitoring; and (3) a single invitation was not sufficient to facilitate adoption of a new mHealth program. Themes were unrelated to participants’ age, race, or ethnicity.

Conclusions: In a study assessing real-world use of mHealth in a group of middle-aged and older adults, we found that—despite free access to the mHealth program. Almost all participants were ≥45 years of age and two-thirds were White. Rapid analysis identified three themes: (1) coaching both facilitates and prevents mHealth adoption and engagement by promoting accountability but leading to guilt among those not meeting goals; (2) preferences regarding the mode of treatment delivery, usability, and treatment content were barriers to mHealth appropriateness and adoption, including preferences for in-person care and a dislike of self-monitoring; and (3) a single invitation was not sufficient to facilitate adoption of a new mHealth program. Themes were unrelated to participants’ age, race, or ethnicity.

Conclusions: In a study assessing real-world use of mHealth in a group of middle-aged and older adults, we found that—despite free access to the mHealth program. Almost all participants were ≥45 years of age and two-thirds were White. Rapid analysis identified three themes: (1) coaching both facilitates and prevents mHealth adoption and engagement by promoting accountability but leading to guilt among those not meeting goals; (2) preferences regarding the mode of treatment delivery, usability, and treatment content were barriers to mHealth appropriateness and adoption, including preferences for in-person care and a dislike of self-monitoring; and (3) a single invitation was not sufficient to facilitate adoption of a new mHealth program. Themes were unrelated to participants’ age, race, or ethnicity.

Keywords: mHealth; implementation; adoption; engagement; weight management; obesity; weight loss; mobile health; veterans; barriers

(JMIR Form Res 2021;5(12):e29916) doi:10.2196/29916

KEYWORDS

mHealth; implementation; adoption; engagement; weight management; obesity; weight loss; mobile health; veterans; barriers
Introduction

Obesity affects roughly 40% of adults in the United States [1] and is associated with many negative outcomes, including social stigma and chronic conditions, such as diabetes and heart disease [2,3]. Although weight management requires long-term, complex behavior change, behavioral weight management programs result in reduced weight, morbidity, and mortality [4-6]. For example, MOVE!, the Veterans Health Administration’s (VHA) flagship weight management program, has been associated with reductions in cardiovascular disease [7] and reductions in diabetes complications [8]. However, while 94% of VHA patients with overweight or obesity are offered weight management programs, only 10% use them [9]. Research outside the VHA in the United States does not typically focus on initial engagement in weight management programs because it is difficult to define the denominator of the target population. However, existing work outside Veterans Affairs (VA) also suggests low engagement rates [10-14].

Mobile health (mHealth) interventions for weight management can result in weight loss outcomes comparable to in-person treatments, although results are mixed [15-17]. For example, Track is an mHealth intervention for adult patients with obesity and related comorbidities (eg, diabetes, hypertension) [18]. It provides app-based self-monitoring in addition to dietician-delivered counseling calls with tailored feedback. In a 12-month effectiveness randomized controlled trial at a community health center, participants randomized to Track were more likely to lose at least 5% of their baseline weight. Given the ubiquity of smartphones [19], increasing access to mHealth for weight management could increase weight management program use and improve patient health. However, there is relatively little information on implementing mHealth for weight management in routine practice settings, particularly among older adults and with regard to patient-level factors [20].

In this work, we performed a qualitative analysis of patient interviews to answer two research questions about implementation outcomes among a sample of primary care patients with obesity: (1) what are barriers and facilitators to mHealth adoption and engagement? and (2) what are patient beliefs about the appropriateness of mHealth for weight management? We conducted this work at the VA Palo Alto Health Care System, which had recently implemented an evidence-based [21] commercial smartphone app designed to facilitate weight management and other health behaviors. Starting in March 2019, the VA Palo Alto Health Care System offered primary care patients with obesity access to an mHealth program for weight management (Vida [21]). The mHealth program was tailored for VHA patients—for example, there was an effort to hire coaches with firsthand military experience or military experience via family members, or those who worked in the military as civilians. Coaches also received online training in military culture via VHA’s Talent Management System.

The program was advertised to and by primary care clinicians, behavioral health staff, and weight management clinicians. It was also advertised at the main hospital’s weekly farmers market, and through flyers and social media. In consultation with the Public Affairs Office, ads and images were selected to represent a diverse population with regard to age, sex, and ethnicity. After learning about the mHealth program in one of the aforementioned ways, patients had to contact VA Palo Alto staff, express interest, and meet minimal criteria (ie, access to a smartphone with internet, no uncontrolled mental or physical health conditions). Patients were then given an access code they could use to download the app and begin the mHealth program.

Between February 2020 and October 2020, we used administrative data to obtain a randomly chosen list of 77 potential participants who were VA Palo Alto adult primary care patients with obesity (ie, body mass index ≥30) who had been given an mHealth access code. We excluded 36 patients from this list who no longer had a BMI ≥30, or who had a hospitalization or suicide attempt in the past 30 days, cognitive impairment, and/or a psychotic disorder diagnosis. We mailed opt-out letters to the remaining 41 patients, calling potential participants who did not opt out or otherwise contact study staff. Interested and eligible patients completed the consent process and a one-time interview over the phone (n=24 for a response rate of 59%). We mailed opt-out letters in February, August, and October 2020. We completed interviews in February, March, June, September, October, and November 2020. Therefore, data collection was complete by November 2020. All participants who completed the consent process also completed the interview. Interviews lasted roughly 45-60 minutes. Detailed, typewritten notes were taken during interviews. COVID-19-related technology problems prevented staff from recording all interviews. However, roughly half of the interviews were digitally audio-recorded (n=11). In most cases, each interview consisted of one staff member and one participant. The semistructured interview guide covered beliefs about weight and weight management and was part of a larger study on that topic. Questions most relevant to the present work are in Table 1 (full guide available upon request). We paid participants $50.
Table 1. Most relevant interview questions and prompts.

<table>
<thead>
<tr>
<th>Question</th>
<th>Relevant prompts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you ever tried to lose weight?</td>
<td>• N/A^a</td>
</tr>
<tr>
<td>Why did you try to lose weight?</td>
<td>• N/A</td>
</tr>
<tr>
<td>How have you tried to lose weight?</td>
<td>• How/where did you learn about/use Vida? • What made it easy to learn about Vida? • What made it difficult to learn about Vida?</td>
</tr>
<tr>
<td>What got in the way of you trying to lose weight?</td>
<td>• What got in the way of using Vida? • Would tailored information have changed your experience with Vida? • Did the coach and examples seem relevant to you? • What would have helped you use Vida? • What do you think of the idea of an app with a live coach to help with weight management?</td>
</tr>
<tr>
<td>What helped you try to lose weight?</td>
<td>• What helped you use Vida? • How did the coach affect your experiences with Vida?</td>
</tr>
</tbody>
</table>

^aN/A: not applicable.

Analysis

We used rapid qualitative analysis methods informed by Neal and colleagues [23] and Nevedal and colleagues [24] to answer the research questions: (1) what are barriers and facilitators to mHealth adoption (initial use) and engagement (continued use, ie, 2 or more uses)? and (2) what are patient beliefs about the appropriateness (ie, perceived fit, relevance, or compatibility [20]) of mHealth for weight management? Rapid analysis has been shown to be faster than and as effective as other forms of qualitative analysis for relatively straightforward research questions such as ours [24,25].

During and after each interview, we completed a matrix in Microsoft Excel (Microsoft Corp) to identify key information about participants’ mHealth experiences. The matrix had a row for each participant. Columns included deductive codes related to our primary research questions: whether participants used Vida, mHealth barriers, and mHealth facilitators. There were societal barriers and facilitators columns to note factors related to sociodemographic information and/or discrimination. There was also an “other notes” column, where researchers could document information that did not fit into the aforementioned columns. After all other information was entered into the matrix, we added columns for participants’ gender, age, and race/ethnicity to help us identify whether there were patterns across those demographic groups. Analysts also wrote a memo for each participant describing the entire interview.

After all interviews were complete, JYB and KA separately performed inductive analysis by reviewing each barriers and facilitators column and looking for themes—that is, repetition of topics, salience of topics to participants, and negative cases (cases that were different or unique compared to other people). They used the same process across rows to identify possible themes within participants, finding none. The authors then met to review themes and resolve discrepancies (there were few), following which the two authors separately reviewed interview memos to search for additional information on existing themes or information on new themes. After meeting to compare results, noting high agreement and no new themes, the authors settled on a final list of themes and representative quotes from participants. All methods were approved by the Stanford University Institutional Review Board.

Results

Overview

We interviewed 24 veterans, 6 of whom were women. Only 2 participants were aged <45 years, and we had roughly equal numbers of participants aged 45-64 years and ≥65 years. Two-thirds of our sample was White; other participants reported their race/ethnicity as Black or African American, Asian, Hispanic, or other. Table 2 provides additional detail. A total of 7 participants reported adopting the mHealth program, 4 of whom were women. For simplicity and to use the language of mHealth, in the sections below we refer to participants who adopted the mHealth program as users and those who did not adopt the mHealth program as nonusers.
We identified three themes related to the adoption of, engagement with, and appropriateness of mHealth, which are described below. Other than noting a greater proportion of women used the mHealth program than men, we identified no themes related to gender, age, or race/ethnicity. Some findings did not fit under the specific themes, but influenced the presentation of themes. First, most participants had tried multiple weight loss methods, including other forms of mHealth. Second, barriers and facilitators to mHealth adoption were similar between users and nonusers, although the latter necessarily described hypothetical reasons. Finally, improving overall health was the main weight management motivator for participants.

**Theme 1: Coaching Both Facilitates and Prevents mHealth Adoption and Engagement**

A participant who used the mHealth program described this theme most succinctly: “the coach…is a double-edged sword, people need accountability, but at the same time they are sometimes afraid of accountability.”

Participants said the live coach facilitated adoption and engagement by facilitating accountability. As a user noted, the coach “makes you responsible to answer to someone,” adding that the accountability was “a big part of [my] success, setting goals and expectations helped me a lot.” This was also true for nonusers, with seven of 17 participants who did not use the mHealth program saying they liked the idea of receiving health coaching via mHealth and two expressing a desire to get another the mHealth program access code after the interview. As one participant who did not use the mHealth program said, the coach was appealing because “I always desire to have someone whom I am accountable to, who is knowledgeable and supportive. In the end we are all humans and need support.” Of note, users and nonusers generally did not report a desire to have a coach matched to them solely on demographic characteristics (“I don’t care if they’re male, female, Black, white”). Instead, they focused on the importance of connecting to the coach in at least one way, which they were as likely to describe as being related to hobbies (eg, types of physical activity) versus a demographic characteristic (eg, gender).

At the same time, participants also described how coaches could prevent adoption and engagement. First, feelings of guilt engendered by not meeting goals led participants to stop meeting with the coach and to stop using the app. For some participants, this was also linked to shame, with one participant stating the following:

“I didn’t like the feeling of answering to someone when I am not successful, I felt like I wasn’t doing enough, and I used to feel ashamed when I would think about talking to the coach. This made me resistant to using Vida.”

For this participant, the feelings applied to doctors as well. For others, the guilt was linked to past health struggles:

“I can’t think of any drug addict or person who has a flaw that likes to continuously focus on it. For me it was like, jeez, I have to keep talking about this. Did I do this food? Did I do this activity? And it’s like ah, I didn’t do it today. I would feel bad because I didn’t stick to the plan. I know I’m not disappointing her, but that’s just the way it would go.”

High coach turnover rates also prevented mHealth engagement as it resulted in some participants having multiple coaches. Participants described frustration arising from multiple rapport-building attempts with new coaches, and ultimately disengagement. As one participant said:

[The coach] enhanced the whole experience, I looked forward to talking to her every week, but then I stopped losing weight which frustrated me…and then I had to travel and my counselor changed. I didn’t feel like building a relationship with a new person.

**Theme 2: Preferences Regarding the Mode of Treatment Delivery, Usability, and Treatment Content**
Were Barriers to mHealth Appropriateness and Adoption

While participants said mHealth was generally appropriate for weight management, they also noted that it was not appropriate in all cases. In this way, the mode of treatment delivery served as a barrier for users and nonusers alike. For example, some participants simply preferred in-person treatment. As a nonuser noted, “I like the convenience of phone meetings, but there is a better accountability when you have in person appointments.” Another nonuser suggested connecting patients to peers, for example, “through a connected app where you’re connected with somebody else in a different part of the country…and you’re both going to lose weight together. You have a friend and you’re committed with somebody.”

Some users described a preference for in-person treatment, despite being willing to try mHealth. One noted that the mHealth program may not have worked for her because the coach “was on the video thing…she wasn’t in my face, she didn’t know me a whole lot,” adding that “it’s a physical thing, in person, not over the phone. I just didn’t connect with [virtual care].”

Half of participants who adopted the mHealth program thought usability was poor—namely, that entering dietary information was more difficult than other apps they used. Some participants noted they were already using a different kind of mHealth for weight management. With regard to treatment content, several participants said they did not use the mHealth program or did not like using the mHealth program because they did not like monitoring their food. Although, one participant noted that “I can’t say that I enjoy logging my food, but I have to acknowledge that it makes a difference.”

Participants also noted that treatment preferences could change with life circumstances and competing demands. For example, some noted that they learned about the mHealth program toward the end of MOVE! and thought it would serve as way to maintain weight management behaviors once MOVE! ended. However, for at least one participant, by the time MOVE! ended, his daughter’s needs superseded his own goals and prevented mHealth adoption. For a user, her husband’s mental health care took a substantial amount of time, which left her limited time to engage in health behavior change. In addition, as another user noted, one’s own health could prevent engagement in mHealth-related behaviors: “I was swimming for a few months and then I had an open sore and I [couldn’t] get into the pool.”

Theme 3: A Single Invitation Was Not Sufficient to Facilitate Adoption of a New mHealth Program

All participants in this study were given an access code for the mHealth program. However, the most common—and therefore primary—barrier to mHealth adoption among participants was not remembering they received an access code, with 12 of 17 nonusers giving this response. Of these 12, two participants remembered seeing ads for the mHealth program, suggesting some familiarity with the program.

Discussion

Principal Findings

This is one of the first studies to assess adoption and appropriateness of mHealth for weight management among a group of middle-aged and older adults in a real-world setting. Most participants viewed mHealth as appropriate for weight management. However, despite seeking out and being given free access to an app with a live coach, the majority did not complete the registration process. As a result, the primary barrier to adoption was not remembering they had access, suggesting potential problems with the feasibility of the implementation of this mHealth intervention. For those who adopted the mHealth intervention, barriers to engagement included guilt and shame related to not meeting goals, competing demands of everyday life, and poor usability. Notably, these barriers were not related to the age of participants. Other than women being more likely to use the mHealth program, we did not find noticeable patterns by gender, age, race, or ethnicity.

Perhaps the most novel finding is that coaching was a double-edged sword for participants. Coaching enticed people to adopt the mHealth intervention. However, if they did not consistently achieve their goals, coaching became a barrier to continued engagement. Both guilt (discomfort with one’s actions) and shame (feeling like a person who cannot achieve goals) were forces for disengagement [26]. There is a literature on guilt and shame in relation to health behaviors. For example, Thøgersen-Ntoumani and colleagues found that higher ratings of self-compassion in response to dietary lapses were associated with less guilt and stronger goal perseverance intentions and self-efficacy during weight loss attempts [27]. These results suggest that promoting self-compassion may encourage participation in coaching interventions when guilt and shame may be at play. However, guilt, shame, and self-compassion have not been a focus of the health coaching literature. Our findings suggest this may be an important area of future research. It is possible that mHealth coaching for weight management requires more advanced training to build the rapport necessary to overcome guilt and shame.

In addition, our findings highlight the importance of acknowledging and working with ambivalence during behavior change interventions. Motivational interviewing, which is a counselling method used to resolve ambivalence, is associated with weight loss in the context of eHealth and telehealth weight loss programs [28]. Although it is a part of Vida coach training, our findings suggest that coaches may need to spend additional time addressing ambivalence to ensure continued engagement. Further, our findings related to guilt, shame, and ambivalence may explain why coaching did not improve outcomes in an mHealth intervention designed to increase physical activity among veterans, even though coaching was front-loaded to increase engagement [29]. These factors should be addressed in future research.

It is also notable that participants generally did not have demographic criteria for coaches. Instead, they focused on a need to connect with the coach across varied affinities. In this study, the mHealth program was adjusted and coaches were...
trained to account for military culture. Therefore, while not
mentioned by participants, it is possible that military-related
tailoring was sufficient for participants to feel an affinity to the
coach. This is heartening given that the intersectional nature of
identity [30] means that it is impossible to demographically match all patients and coaches. At the same time, data from
in-person medical settings suggest that demographic matching
can influence health outcomes; for example, mortality is lower
for Black infants when Black physicians provide care [31]. An
important area of future research includes determining whether
outcomes are better when coaches are matched on specific characteristics and whether this is more important for historically
marginalized populations.

Our results also add to other findings from VHA patients using
a web-based weight management program. A qualitative study
of women veterans using an online version of the Diabetes
Prevention Program had some similar results—the women liked
the ability to access intervention materials at any time on the
internet, but some did not like self-monitoring [32]. That digital
program resulted in higher participation rates than standard
VHA weight management programming [33], further highlighting the potential utility of digital health in this context.

Our most actionable finding may be that a single referral to
mHealth is insufficient for adoption. Although this finding is
related to a relatively simple theme, solutions may be complex.
The Expert Recommendations for Implementing Change (ERIC)
suggest implementation strategies that might facilitate adoption,
including developing and distributing educational materials,
obtaining and using patient and family feedback, and using mass
media [34]. In the context of mHealth, this could include repeated follow-up contacts via phone, SMS text message–based
reminders, and/or marketing materials sent to interested patients
more than once. Implementation efforts of other digital health
programs in VHA suggest that engaging clinicians and
leadership will be especially important [35]. Health care systems
could also learn from the private sector while using resources
like the Digital Health Checklist to ensure a match between
private sector and health care ethical standards [36].

Limitations

Limitations include a focus on VHA participants. VHA is an
integrated health care system that provides primary and specialty
physical and mental health care to its patients. People receiving
care in stand-alone clinics may have different experiences. In
addition, we did not have objective information on whether
participants used a specific mHealth program within the app
(eg, a program that had a specific number of sessions) as barriers
and facilitators could differ across finite versus infinite programs. We also do not have information regarding
participants’ views of military-related tailoring. Although the
random sample of users invited to participate in this research
should have helped account for potential differences between
people who used the program earlier versus later, given our
small sample and limited information about the denominator
of people offered the program, the sample may not be
representative.

The primary limitation of this study relates to a primary
collection—few participants who received an access code to
a new mHealth program adopted it. Including fewer users than
nonusers in our sample could have led to missed themes related
to adoption. However, similarities in descriptions of barriers
and facilitators to adoption between users and nonusers
ameliorate this concern. Other strengths include our older sample
and an analysis based on real-world mHealth use, which are not
commonly studied.

Conclusions

Our findings suggest that implementing an mHealth intervention
for weight management in an integrated system primarily
serving older adults requires more than one information session.
Findings also suggest that focusing future research on the
coaching relationship and how users’ lives and goals change
over time will be important for facilitating engagement and
improved health. Most participants thought the mHealth
intervention was appropriate for weight management, with some
nevertheless preferring in-person care. Therefore, the best way
to ensure equitable care will be to ensure multiple routes to
achieving the same behavioral health goals. VHA patients have
the option of using mHealth for weight management, but can
also attend group weight management programs or
single-session nutrition classes, or access fitness facilities.
Health care policy does not allow this for most people in the
United States; however, expanded access to behavioral weight
management programs is an important long-term goal to ensure
health for all.

Acknowledgments

The authors would like to thank Dr Andrea Nevedal for her methodological insight. The views are those of the authors and do
not necessarily represent those of the Department of Veterans Affairs (VA) or the United States government. JYB is supported
by a VA Health Services Research & Development Career Development Award (CDA 15-257) at the VA Palo Alto. This work
was supported by the Research in Implementation Science for Equity (RISE) program funded by the National Heart, Lung, and
Blood Institute (5R25HL126146-05) through the Programs to Increase Diversity Among Individuals Engaged in Health-Related
Research (PRIDE).

Authors’ Contributions

JYB conceptualized the study, collected data, conducted analysis, and led writing of the manuscript. KA collected data, conducted
analysis, and contributed to writing the manuscript. RM contributed to writing the manuscript.
References


Abbreviations

ERIC: Expert Recommendations for Implementing Change
mHealth: mobile health
VA: Veterans Affairs
VHA: Veterans Health Administration

Eman M Alanazi1,2, MSc; Aalaa Abdou3, MD; Jake Luo4, PhD

1Department of Health Informatics, College of Health Sciences, Saudi Electronic University, Riyadh, Saudi Arabia
2Department of Biomedical and Health Informatics, College of Engineering, University of Wisconsin-Milwaukee, Milwaukee, WI, United States
3Department of Radiotherapy, Children's Cancer Hospital 57357, Cairo, Egypt
4Department of Health Informatics and Administration, College of Health Sciences, University of Wisconsin-Milwaukee, Milwaukee, WI, United States

Corresponding Author:
Eman M Alanazi, MSc
Department of Health Informatics
College of Health Sciences
Saudi Electronic University
Abi Bakr As Siddiq Branch Rd
Riyadh, 13323
Saudi Arabia
Phone: 966 112613500
Email: e.alanazi@seu.edu.sa

Abstract

Background: Stroke, a cerebrovascular disease, is one of the major causes of death. It causes significant health and financial burdens for both patients and health care systems. One of the important risk factors for stroke is health-related behavior, which is becoming an increasingly important focus of prevention. Many machine learning models have been built to predict the risk of stroke or to automatically diagnose stroke, using predictors such as lifestyle factors or radiological imaging. However, there have been no models built using data from lab tests.

Objective: The aim of this study was to apply computational methods using machine learning techniques to predict stroke from lab test data.

Methods: We used the National Health and Nutrition Examination Survey data sets with three different data selection methods (ie, without data resampling, with data imputation, and with data resampling) to develop predictive models. We used four machine learning classifiers and six performance measures to evaluate the performance of the models.

Results: We found that accurate and sensitive machine learning models can be created to predict stroke from lab test data. Our results show that the data resampling approach performed the best compared to the other two data selection techniques. Prediction with the random forest algorithm, which was the best algorithm tested, achieved an accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and area under the curve of 0.96, 0.97, 0.96, 0.75, 0.99, and 0.97, respectively, when all of the attributes were used.

Conclusions: The predictive model, built using data from lab tests, was easy to use and had high accuracy. In future studies, we aim to use data that reflect different types of stroke and to explore the data to build a prediction model for each type.

(JMIR Form Res 2021;5(12):e23440) doi:10.2196/23440

KEYWORDS
stroke; lab tests; machine learning technology; predictive analytics

Introduction

Stroke is a neurological deficit, primarily because of acute central nervous system focal injury caused by a vascular issue. It is a major cause of disability and death worldwide [1]. It estimated that the overall prevalence of stroke in the United States is 2.5%, and about 7 million Americans over the age of 20 years have experienced a stroke. The condition has a significant negative impact on patients' health and quality of life. It also has a negative impact on hospital services and the availability of beds and was estimated to cost the US economy about US $351.2 billion between 2014 and 2015 [2]. There are
two types of stroke: ischemic and hemorrhagic. Hemorrhagic stroke occurs because of a burst vessel that leads to bleeding in the brain, whereas ischemic stroke occurs because of a blockage of the arteries of the brain. Ischemic strokes are the most common, comprising 85% to 90% of all strokes [3]. This condition can be prevented by promoting health and increasing awareness of risk factors. There are many risk factors related to lifestyle, including obesity, diet, alcohol intake, and lack of physical activity [4]. Underlying conditions, such as diabetes, hypertension, and cardiovascular diseases, may also lead to stroke. Therefore, proper self-management of these diseases and the pursuit of a healthy lifestyle may prevent the occurrence of stroke.

In 2019, the American College of Cardiology/American Heart Association released the Guideline on the Primary Prevention of Cardiovascular Disease. The guideline recommends a complete assessment and examination of patients who are at risk of developing blockages in their arteries that may lead to a heart attack or stroke and might die as a result [5]. Now more than ever, physicians can access clinical evidence to identify high-risk patients using approaches such as acquiring a complete patient history and conducting thorough physical exams for risk assessment. Patient records contain many useful predictive factors, such as patient demographic (eg, age and gender), lifestyle (eg, diet and physical activity), and existing medical condition factors (eg, diabetes and hypertension), that might lead to stroke [5]. The growth of arterial blockages and decades of damage to blood vessels, which may lead to stroke, are often associated with these risk factors. If physicians can assess the risks of stroke easily and conveniently, strokes could be prevented at an earlier stage. This approach could save lives and reduce the economic burden of health care services. In the age of artificial intelligence and machine learning, a clinical decision support system has been developed to assist physicians to diagnose and identify individuals with a high risk of stroke. The potential of applying machine learning technologies in the cardiovascular domain is significant, from identifying individuals with a high risk of stroke [6,7] to predicting outcomes of patients following treatment [8,9]. Most of these studies use either health habits and lifestyle factors, such as smoking or alcohol consumption; conditions that predispose to strokes, such as hypertension and diabetes mellitus; or neuroimaging, such as computed tomography and magnetic resonance imaging, to either classify or predict the disease.

Besides assessing known risk factors for stroke, scientists are trying to develop lab tests that can predict stroke. One of the major advantages of using lab test results for prediction is that lab tests are commonly collected in clinical settings, and the information is often well documented in patients’ records. In this study, we explored data-driven approaches using supervised machine learning models to predict the risk of stroke from different lab tests.

Several studies have been able to identify independent laboratory tests that are correlated with stroke using descriptive statistical analysis. Sughrue et al [10] conducted a retrospective study in 2013 that identified 35 tests with a statistically significant correlation with a future stroke diagnosis. The most informative tests were for various types of cholesterol. Two of these 35 laboratory tests were urine tests, and 33 were blood, serum, or plasma tests. Some tests were positively associated with an outcome of stroke (ie, neutrophil count and percent; CD3+, CD8+, and T8 suppressor cells; monocytes; eosinophils; and CD3 cells), and others were negatively correlated (ie, hematocrit and lymphocytes). Their results show that it is possible to correlate future stroke with collected lab test data. Farah and Samra [11] conducted a retrospective study investigating the association between the neutrophil-to-lymphocyte ratio (NLR), mean platelet volume (MPV), and the risk of stroke. Two-tailed t tests showed no significant differences in the stroke group’s MPV values compared with those in the control group. However, the NLRs of the stroke patients were significantly different compared with those of the control group. That study indicated the existence of a correlation between the level of NLR and stroke risk. NLR levels have been shown to be higher in stroke patients than in control groups. Feng et al [12] reviewed the scientific literature on the potential role and the possible epidemiological relationships between red cell distribution width (RDW) and ischemic stroke in a meta-analysis of 40 manuscripts from China National Knowledge Infrastructure and PubMed databases. They reported that patients with stroke had higher levels of RDW than those without strokes. Another study by Kaya et al [13] also investigated the association between baseline RDW level and stroke risk in patients with heart failure. These authors found that heart failure patients suffering from stroke had significantly increased basal RDW levels (mean 16.9, SD 1.4, vs mean 14.8, SD 1.6; P<.001) and serum uric acid levels (mean 8.8, SD 1.7, vs mean 7.5, SD 1.1; P=.027) compared with patients without stroke, according to the propensity score analysis. Giles et al [14] used data from a national cohort to investigate whether low folate levels were associated with ischemic stroke and found that folate concentrations of ≤9.2 nmol/L could be a risk factor for ischemic stroke (relative risk 1.37, 95% CI 0.82-2.29). Another study by Qin et al [15] concluded that there is a significant risk of first ischemic stroke in hypertensive patients with low levels of folate and vitamin B12.

These studies demonstrate the value of lab test results for predicting stroke. Our study aimed to leverage lab test results to build machine learning models for stroke prediction. We prepared the data sets using three data selection techniques for this study. After that, for each data selection technique, we applied four individual machine learning classifiers to prepare prediction models. We measured the performance of each prediction model using six different performance measures. Our results indicate that the data resampling technique outperformed the decision tree and random forest classifiers.

Methods

Overview

Figure 1 shows the outline of our investigation. In the first step, we collected data from the National Health and Nutrition Examination Survey (NHANES). In the second step, we selected the data using three data techniques for our prediction models. The first one was conducted without data resampling, the second

https://iformative.jmir.org/2021/12/e23440

JMIR Form Res 2021 | vol. 5 | iss. 12 | e23440 | p.224

(page number not for citation purposes)
one included data imputation, and the third one was conducted with data resampling.

We used 10-fold cross-validation to perform the train and test approach. To train models, we used four different machine learning classifiers, and six performance measures were used to assess the performance of the models. The elaborated descriptions of the data sets, classifiers, and performance metrics that were used are given below.

**Figure 1.** Flow diagram of the study methodology. NHANES: National Health and Nutrition Examination Survey.

### Data Collection

The NHANES survey was conducted to examine the health and nutritional status of adults and children in the United States; “NHANES is a major program of the National Center for Health Statistics (NCHS). NCHS is part of the Centers for Disease Control and Prevention (CDC) and has the responsibility for producing vital and health statistics for the Nation” [16]. The data sets contain five domains: demographics, dietary data, examination data, laboratory data, and questioner data. Each domain contains several subdomains. Our focus was on data sets that contain information about laboratory tests. The data sets are available from 1999 to 2017, and we considered data from 2011 to 2015. The total number of participants was 15,714 during this period. To reduce the impact of imbalanced data, we noted that in the entire data set, there were about 17% of participants who had experienced a stroke. Therefore, we included total of 4186 participants, of whom 608 (14.5%) had experienced a stroke (Figure 2). The list of data attributes is shown in Table 1. The data sets contained 21 attributes, including each patient’s age and gender as well as other lab test information for each respective patient. The data sets and their information are available online [16], where the data are presented from the year 2000 to the current year. For this study, the data were collected for each year and combined using the sequence number (SEQN). After combining and cleaning the data, we used the Waikato Environment for Knowledge Analysis (WEKA; version 3.8.0) system to build and test machine learning models.
Figure 2. Participant selection and prevalence of stroke in the National Health and Nutrition Examination Survey (NHANES).

Table 1. List of the data attributes.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Years</td>
</tr>
<tr>
<td>Gender</td>
<td>N/A b</td>
</tr>
<tr>
<td>Albumin, urine</td>
<td>ug/mL</td>
</tr>
<tr>
<td>Creatinine, urine</td>
<td>mg/dL</td>
</tr>
<tr>
<td>White blood cell count</td>
<td>1000 cells/μL</td>
</tr>
<tr>
<td>Lymphocytes</td>
<td>1000 cells/μL</td>
</tr>
<tr>
<td>Monocytes</td>
<td>1000 cells/μL</td>
</tr>
<tr>
<td>Segmented neutrophils</td>
<td>1000 cells/μL</td>
</tr>
<tr>
<td>Eosinophils</td>
<td>1000 cells/μL</td>
</tr>
<tr>
<td>Basophils</td>
<td>1000 cells/μL</td>
</tr>
<tr>
<td>Red blood cell count</td>
<td>Million cells/μL</td>
</tr>
<tr>
<td>Hemoglobin</td>
<td>g/dL</td>
</tr>
<tr>
<td>Hematocrit</td>
<td>%</td>
</tr>
<tr>
<td>Mean cell volume</td>
<td>fl</td>
</tr>
<tr>
<td>Mean cell hemoglobin</td>
<td>pg</td>
</tr>
<tr>
<td>Mean corpuscular hemoglobin concentration</td>
<td>g/dL</td>
</tr>
<tr>
<td>Red cell distribution width</td>
<td>%</td>
</tr>
<tr>
<td>Platelet count</td>
<td>1000 cells/μL</td>
</tr>
<tr>
<td>Mean platelet volume</td>
<td>fl</td>
</tr>
<tr>
<td>Cotinine, serum</td>
<td>ng/mL</td>
</tr>
<tr>
<td>Red blood cell folate</td>
<td>mg/dL</td>
</tr>
</tbody>
</table>

Classification

Several different machine learning algorithms can handle a binary classification problem. In this study, we used four machine learning algorithms: naïve Bayes, BayesNet, J48 (Java implementation of C4.5 algorithm), and random forest. The performance of the algorithms was evaluated and compared for

---

*All data types were numeric, except for ‘gender,’ which was nominal.

bN/A: not applicable; this type of data did not have units.*
stroke prediction using lab test results as features. Details of the algorithms are as follows:

- The J48 algorithm creates a tree based on the C4.5 algorithm with pruning.
- The random forest algorithm creates a forest of random trees and outputs the mode of the classes created by individual trees.
- The naïve Bayes algorithm creates a classifier based on the naïve Bayes method, which assumes that all attributes are independent.
- The BayesNet algorithm creates a classifier based on non-naïve Bayes, which does not assume that all attributes are independent.

In the cross-validation approach, the data sets are divided into several equal portions; in general, 5-fold and 10-fold cross-validations are used when the data sets are equally divided into 5 and 10 portions [17]. With this approach, for each simulation, one portion of each data set is used to train the prediction model and the rest are used for validation. In this study, we used 10-fold cross-validation and, in this process, we divided the whole of each data set into 10 equal parts; each time, 10% of each data set was used to train the model and 90% was used for validation. In this task, three data analyses were conducted where the first data analysis applied each of the machine learning techniques on the data sets without data manipulation or resampling. The aim was to determine the baseline for the data sets among the various machine learning techniques. The imputation of missing data set entries was conducted in the second analysis. In statistics, imputation entails substituting missing data with values calculated using any of a number of techniques [18]. Imputation is a useful technique in remedying missing data, since missing data may lead to inaccurate predictions. We used the default ReplaceMissingValue filter in WEKA, which replaces all missing values for nominal and numeric attributes in a data set with the means and modes from the training data. Most of the features had 5% missing values, and one feature had 11% missing values. After the imputation of the missing data, data resampling was conducted in the third analysis. Data resampling is a commonly used technique, since training may result in nonuniformity of class labels. In this case, the resampling technique was applied to select a specific subset of data points for model training [19]. After resampling the data, the results of the first analysis should be improved because of the balancing of the data set distribution. A balanced distribution was achieved through the use of WEKA, which randomly resamples the data. Based on the available theoretical knowledge about resampling and imputation in statistics, the results after the third analysis should be improved.

**Evaluation Metrics**

Model accuracy was evaluated based on the following measures: recall or sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and area under the curve (AUC) (or area under the receiver operating characteristic [ROC] curve) to compare the four classifiers. Details of these measures are as follows:

- **Sensitivity**, also known as recall or true positive rate, is the number of true positives divided by the number of true positives plus the number of false negatives. It is the likelihood that the patient has a high risk of stroke [20].
- **Specificity**, also known as the true negative rate, is the proportion of individuals classified as non-stroke to the total number of actual non-stroke cases. It is the likelihood that a patient who does not have a risk of stroke will have a negative result [21].
- **PPV**, also known as precision, is the number of true positives divided by the number of true positives plus the number of false positives. It is the proportion of individuals who have suffered a stroke to the total number of participants classified as having a risk of stroke [22].
- **NPV** is the percentage of negative tests in patients who are free from the disease or the proportion of individuals who have not suffered a stroke to the total number of participants classified as not having a risk of stroke [22].
- **Overall accuracy** is the number of correctly classified instances over the total size of the data set [20].
- **The AUC** is the area under the ROC curve, which is constructed by plotting the true positive rate against the true negative rate [23].

We will also look at the Pearson correlation coefficient value of each independent predictor to investigate the relationship between each lab test and risk of stroke.

**Results**

In the NHANES data sets, 608 participants suffered from a stroke from 2011 to 2015. The median age of participants who had a stroke was 51 years for both men and women. The numbers of men and women who had a stroke were 220 (36.2%) and 190 (31.3%), respectively; 198 (32.6%) participants did not reveal their gender identity.

After the data collection process, the data were analyzed in three ways: without data resampling, with data imputation, and with data resampling. Data resampling techniques were used to tackle data imbalance problems in the data sets. These sampling techniques are widely used in machine learning–based prediction models in different areas [24]. Our first analysis was done without the data resampling technique, where the four machine learning algorithms were applied directly to the data sets. The first analysis produced poor results for all four classifiers. The best sensitivity rate among the classifiers in the first analysis was for the BayesNet model, followed by the naïve Bayes model.

In the second analysis, we applied the data imputation technique to the data sets, which replaced missing values and deleted features that had more than 50% missing values; the prediction accuracy improved for all models, except for the naïve Bayes model, whose performance decreased slightly after replacing the missing values.

In the third analysis, we resampled the data. After resampling, the prediction accuracy improved significantly for both the decision tree and random forest models, but only slightly for the naïve Bayes and BayesNet models. Table 2 shows the scores of accuracy, sensitivity, specificity, PPV, NPV, and AUC, according to the three data analysis techniques and four
classifiers. The table shows that the random forest model was the best classifier with the data resampling technique. Figures 3 and 4 show the score comparisons among the three data selection techniques for the decision tree and random forest models, respectively. We considered the decision tree and random forest classifiers to compare the performance, as they significantly improved the performance in the third analysis. Both figures clearly show that the third analysis, the data resampling technique, outperformed the other two techniques for the decision tree and random forest classifiers.

Table 2. Results of three data analysis techniques.

<table>
<thead>
<tr>
<th>Technique and classifier</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV(^a)</th>
<th>NPV(^b)</th>
<th>AUC(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Without data resampling</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>0.82</td>
<td>0.34</td>
<td>0.88</td>
<td>0.27</td>
<td>0.91</td>
<td>0.76</td>
</tr>
<tr>
<td>BayesNet</td>
<td>0.82</td>
<td>0.38</td>
<td>0.89</td>
<td>0.37</td>
<td>0.90</td>
<td>0.88</td>
</tr>
<tr>
<td>Decision tree</td>
<td>0.83</td>
<td>0.33</td>
<td>0.87</td>
<td>0.14</td>
<td>0.95</td>
<td>0.73</td>
</tr>
<tr>
<td>Random forest</td>
<td>0.86</td>
<td>0.55</td>
<td>0.86</td>
<td>0.01</td>
<td>0.99</td>
<td>0.87</td>
</tr>
<tr>
<td><strong>Data imputation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>0.81</td>
<td>0.32</td>
<td>0.88</td>
<td>0.25</td>
<td>0.91</td>
<td>0.74</td>
</tr>
<tr>
<td>BayesNet</td>
<td>0.86</td>
<td>0.53</td>
<td>0.92</td>
<td>0.54</td>
<td>0.92</td>
<td>0.85</td>
</tr>
<tr>
<td>Decision tree</td>
<td>0.88</td>
<td>0.61</td>
<td>0.91</td>
<td>0.46</td>
<td>0.95</td>
<td>0.74</td>
</tr>
<tr>
<td>Random forest</td>
<td>0.90</td>
<td>0.89</td>
<td>0.90</td>
<td>0.33</td>
<td>0.99</td>
<td>0.85</td>
</tr>
<tr>
<td><strong>Data resampling</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>0.82</td>
<td>0.33</td>
<td>0.88</td>
<td>0.29</td>
<td>0.90</td>
<td>0.74</td>
</tr>
<tr>
<td>BayesNet</td>
<td>0.87</td>
<td>0.53</td>
<td>0.93</td>
<td>0.57</td>
<td>0.92</td>
<td>0.85</td>
</tr>
<tr>
<td>Decision tree</td>
<td>0.93</td>
<td>0.76</td>
<td>0.95</td>
<td>0.72</td>
<td>0.96</td>
<td>0.86</td>
</tr>
<tr>
<td>Random forest</td>
<td>0.96</td>
<td>0.97</td>
<td>0.96</td>
<td>0.75</td>
<td>0.99</td>
<td>0.97</td>
</tr>
</tbody>
</table>

\(^{a}\)PPV: positive predictive value.

\(^{b}\)NPV: negative predictive value.

\(^{c}\)AUC: area under the curve.

Figure 3. Performance comparison among three data selection techniques for the decision tree model. AUC: area under the curve; NPV: negative predictive value; PPV: positive predictive value.
Table 3 shows the results from Pearson correlation analysis of the independent predictors.

Table 3. Pearson correlation coefficient values of independent predictors.

<table>
<thead>
<tr>
<th>Independent predictor of stroke</th>
<th>Pearson correlation coefficient (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.26</td>
</tr>
<tr>
<td>Gender</td>
<td>0.13</td>
</tr>
<tr>
<td>Red cell distribution width (%)</td>
<td>0.18</td>
</tr>
<tr>
<td>Lymphocytes (%)</td>
<td>0.15</td>
</tr>
<tr>
<td>Red blood cell folate (ng/mL)</td>
<td>0.13</td>
</tr>
<tr>
<td>Segmented neutrophils (%)</td>
<td>0.12</td>
</tr>
<tr>
<td>Hemoglobin (g/dL)</td>
<td>0.11</td>
</tr>
<tr>
<td>Red blood cell count (million cells/μL)</td>
<td>0.11</td>
</tr>
<tr>
<td>Hematocrit (%)</td>
<td>0.09</td>
</tr>
<tr>
<td>Lymphocytes (1000 cells/μL)</td>
<td>0.08</td>
</tr>
<tr>
<td>Segmented neutrophils (1000 cell/μL)</td>
<td>0.07</td>
</tr>
</tbody>
</table>

**Discussion**

**Principal Findings**

From the previous section, we noticed that our models had the potential to perform stroke prediction using lab test data. Our results show that the random forest model was the best classifier after conducting the data resampling technique.

Also, several observations can be made from the results in Table 3. We identified nine lab tests, in addition to age and gender, that effectively correlated with stroke occurrence. These correlations were calculated using the Pearson correlation coefficient. These results align with other research that showed a linear relationship between some of these variables and stroke. Several studies have shown that age is correlated with the risk of stroke. According to Muntner et al [2], stroke incidence doubles after the age of 45 years, and 70% of all strokes occur over the age of 65 years. Many studies have investigated the relationship between baseline RDW and stroke. They found that elevated RDW is a risk factor in ischemic stroke [12,13,25].

One of the novel correlations that were found in this study is the lymphocyte percentage. Lymphocytes are white blood cells, including B cells, T cells, and natural killer cells. Lymphocyte percentage is positively associated with stroke occurrence. There have been no studies suggesting that lymphocyte percentage can be a predictor of stroke, but different studies have examined the use of immune cells as biomarkers to predict stroke outcome [26,27]. There is one study that showed a negative correlation between hematocrit and stroke occurrence [10]. Folate deficiency has various clinical manifestations. Our finding that serum folate level was correlated with the risk of stroke is in line with the finding of Giles et al [14], who found that a serum folate concentration of ≤9.2 nmol/L may slightly increase the risk for ischemic stroke. Other studies have shown that folic acid therapy involving folic acid, vitamin B12, and vitamin B6 reduced the risk of ischemic stroke [15,28]. Neutrophils, which are normally the most abundant circulating white blood cells and respond quickly to infection, also contribute to the main processes causing an ischemic stroke, as they facilitate the development of blood clots. Neutrophils are, therefore, also of...
considerable importance as targets for treating and preventing ischemic stroke [29]. A study by Sughrue et al [10] produced results similar to ours regarding the positive association between neutrophils and stroke occurrence. Hemoglobin levels can predict the risk of stroke. Observational studies have reported an independent association between red blood cell count, hematocrit, and hemoglobin concentration and the risk of developing stroke [30,31].

The correlations between these different lab tests and stroke were found in several studies. However, this is the first study that used all of these different attributes to build a prediction model using machine learning algorithms. Our results showed that a prediction model can be created using the random forest algorithm and could achieve an accuracy of 0.96.

Conclusions

Machine learning applications are becoming more widely used in the health care sector. The prediction of stroke using machine learning algorithms has been studied extensively. However, no previous work has explored the prediction of stroke using lab tests. The results of several laboratory tests are correlated with stroke. Building a prediction model that can predict the risk of stroke from lab test data could save lives. In this study, we created a prediction model using the random forest algorithm and achieved a 96% accuracy rate. The model can be integrated with electronic health records to provide a real-time prediction of stroke from lab tests. Because of the nature of the data, we could not predict the type of stroke: hemorrhagic or ischemic. In future studies, we aim to use data that provide information about different types of stroke to build prediction models for each type.

Acknowledgments

EMA conducted the research design, data collection, and data analysis and wrote the original draft. AA assisted with the literature review of the lab tests. JL revised and edited the original draft and provided guidance throughout the whole research process. This study received no external funding.

Conflicts of Interest

None declared.

References


Abbreviations

AUC: area under the curve  
CDC: Centers for Disease Control and Prevention  
MPV: mean platelet volume  
NCHS: National Center for Health Statistics  
NHANES: National Health and Nutrition Examination Survey  
NLR: neutrophil-to-lymphocyte ratio  
NPV: negative predictive value  
PPV: positive predictive value  
RDW: red cell distribution width  
ROC: receiver operating characteristic  
SEQN: sequence number  
WEKA: Waikato Environment for Knowledge Analysis
Using Artificial Neural Network Condensation to Facilitate Adaptation of Machine Learning in Medical Settings by Reducing Computational Burden: Model Design and Evaluation Study

Dianbo Liu¹,², PhD; Ming Zheng², BSc; Nestor Andres Sepulveda², PhD

¹Department of Biomedical Informatics, Harvard University, Boston, MA, United States
²Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, United States

Abstract

Background: Machine learning applications in the health care domain can have a great impact on people’s lives. At the same time, medical data is usually big, requiring a significant number of computational resources. Although this might not be a problem for the wide adoption of machine learning tools in high-income countries, the availability of computational resources can be limited in low-income countries and on mobile devices. This can limit many people from benefiting from the advancement in machine learning applications in the field of health care.

Objective: In this study, we explore three methods to increase the computational efficiency and reduce model sizes of either recurrent neural networks (RNNs) or feedforward deep neural networks (DNNs) without compromising their accuracy.

Methods: We used inpatient mortality prediction as our case analysis upon review of an intensive care unit dataset. We reduced the size of RNN and DNN by applying pruning of “unused” neurons. Additionally, we modified the RNN structure by adding a hidden layer to the RNN cell but reducing the total number of recurrent layers to accomplish a reduction of the total parameters used in the network. Finally, we implemented quantization on DNN by forcing the weights to be 8 bits instead of 32 bits.

Results: We found that all methods increased implementation efficiency, including training speed, memory size, and inference speed, without reducing the accuracy of mortality prediction.

Conclusions: Our findings suggest that neural network condensation allows for the implementation of sophisticated neural network algorithms on devices with lower computational resources.

(JMIR Form Res 2021;5(12):e20767) doi:10.2196/20767

KEYWORDS

artificial neural network; electronic medical records; parameter pruning; machine learning; computational burden;
medical machine learning application wherein the mortality of a patient after being transferred to the intensive care unit (ICU) can be predicted based on their vital signs, laboratory tests, demographics, and other factors. Mortality prediction is important in clinical settings because such a prediction can help determine the declining state and need for intervention. We built baseline models with either recurrent neural network (RNN) or dense neural network architectures, based on which we explored efficiency improvements via neural network condensation without sacrificing the prediction accuracy. An RNN is a class of artificial neural networks wherein connections between nodes form a directed graph along a temporal sequence that consider a sequence of input in a recurrent manner. RNNs are widely used in clinical informatics in tasks such as temporal data analysis and clinical natural language processing.

Reduction of complexity and improvement of efficiency of artificial neural networks is an active field of research, wherein a wide range of methods have been explored. One representative example is neural network pruning, wherein a fraction of weights is removed from the trained model and the “lottery ticket” is found when the remaining weight can still be quickly trained with competitive loss and accuracy [1-3]. There are more fancy pruning approaches where the authors use another neural network to learn and conduct the best pruning decisions considering the network to be pruned (ie, the backbone neural network). For example, Lin et al [4] developed a method called runtime neural pruning to model their pruning process as a Markov decision process and use reinforcement learning for training via an additional RNN. Zhong et al [5], on the other hand, used long short-term memory (LSTM) to guide an end-to-end pruning of the backbone neural network. Some other previous works have converted the neural network condensation into an optimization problem where parameters are penalized under some norm [6-9]. One RNN-specific condensation method is that instead of embedding information into multiple recurrent layers, we only use one recurrent layer but extend the capacity of the RNN unit (cell) by incorporating more hidden layers within the cell. Dai et al [10] showed that DNNs were inserted between the recurrent layer and the input (masking) layer for each gate in the LSTM to form an LSTM embedded with hidden layers (ie, hLSTM). Such an architecture can, in principle, be more efficient (ie, fewer number of parameters and higher training speed). There is another posttraining condensation method called *quantization*, wherein parameters originally stored in a 32-bit floating point format are forcibly converted to 8 fixed bits [11]. Other methods used for neural network condensation include, but are not limited to, binarization of neural networks [12], knowledge distillation [13], and Huffman coding [14]. In this paper, we describe the use of hLSTM, neural network pruning, and quantization to condense the size of neural networks and increase speed while maintaining their prediction accuracy.

**Methods**

### Intensive Care Unit Data

We used the Medical Information Mart for Intensive Care-III (MIMIC-III) critical care database for the implementation of our models [15]. In all, 53,423 distinct hospital adult patients admitted to critical care units between 2001 and 2012 are included in this database. We excluded all neonatal and pediatric patients (aged 18 years or younger at the time of ICU stay) because the physiology of pediatric critical care patients differs significantly from that of adults [16]. We also excluded any hospital admissions with multiple ICU stays or transfers between different ICU units. The final cohort comprised 33,798 unique patients, with a total of 42,276 hospital admissions and ICU stays. Of these 33,798 patients, we defined a test set of 5070 (15%) patient stays. In-hospital mortality was determined by comparing patient date of death with hospital admission and discharge times. The mortality rate within the cohort was 10.9%. The median age of adult patients was 65.8 (SD 11.3) years, and 55.9% (18,893/33,798) patients were male. A mean of 4579 (SD 721.7) charted observations and 380 (SD 215.8) laboratory measurements, as well as other static information, are available for each hospital admission.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality during ICU³ stay, n (%)</td>
<td>3717 (10.9)</td>
</tr>
<tr>
<td>Age in years, median (SD)</td>
<td>65.8 (11.3)</td>
</tr>
<tr>
<td>Male participants</td>
<td>18,893 (55.9)</td>
</tr>
</tbody>
</table>

³ICU: intensive care unit.

### Data Prepossessing

Data were collected from the MIMIC-III database. Only data from the first 48 hours were used as inputs in our analysis. For the purpose of this study, 76 features were selected for analysis (see examples listed in *Textbox 1*). Some features may appear multiple times (in different means or conditions) and are thus regarded as *independent* features. We resampled the time series into regularly spaced intervals. If there were multiple measurements of the same variable in the same interval, we used the value of the last measurement. We imputed the missing values using the previous value, if it exists, or a prespecified “normal” [16] value, otherwise. In addition, we added a binary mask input for each variable, indicating the time steps that contain a true (vs imputed) measurement [17]. Categorical variables were encoded using a one-hot vector at each time step. Then, the inputs were normalized by subtracting the mean and dividing it by the SD value. Statistics were calculated per variable after imputation of missing values.
Textbox 1. Examples of the 76 features selected for the analysis.

- pH
- Fraction-inspired oxygen
- Systolic blood pressure
- Height
- Weight
- Oxygen saturation
- Diastolic blood pressure
- Glucose
- Temperature
- Mean blood pressure
- Capillary refill rate
- Respiratory rate
- Heart rate
- Fraction inspired oxygen
- Glasgow Coma Scale–50

Performance Metrics
Classification accuracy of all models were measured using area under the receiver operating curve AUROC (also called AUCROC) on the test set. Sizes of model were measured by the number of parameters and sizes of the saved model file. Inference speed was calculated based on time taken to make predictions on test data and was normalized per patient. We used Python 3.6, Keras 2.2.4 with TensorFlow 1.1.2, as the backend for the analysis.

RNN Model
Our RNN baseline model is designed as an RNN consisting of a masking layer, two LSTM layers, a dropout layer, and a dense output layer, as shown in Figure 1. We chose two layers of LSTM because, based on a literature review, we identified this structure to be the one with the best performance in the MIMIC-III mortality prediction work [16]. The masking layer masks (skips) the time step for all downstream layers if the values of input tensor at the time step are all equal to zero, which represents missing data for that time step. The first layer of LSTM takes in the original 76 features and generates a 16-feature hidden state based on the hidden state of the previous step and the new incoming observation. Then, such a hidden state is forward to the entrance of the second LSTM layer, which produces another 16-feature hidden state at each step. A dropout layer is followed by the last-step hidden state of the second LSTM layer to prevent complex coadaptations of the neurons. Finally, a dense layer is used to generate a soft 0/1 mortality prediction. The training was conducted using Adam algorithm with a dropout rate of 0.3 between layers and a learn rate of 0.001. In this study, hyperparameters were chosen by grid searching based on performance on the validation set.
Figure 1. Architecture of recurrent neural network baseline model. DNN: deep neural network; LSTM: long short-term memory; ReLU: rectified linear unit.

**hLSTM Model**
Besides pruning upon RNN, we also tried another way by inserting an additional hidden dense layer into the inner gates of LSTM, which we called hLSTM, to improve the “power” of the LSTM. For a traditional LSTM, the inner structure is as follows:

where * is the matrix product; \( \otimes \) is the element-wise product; \( W \) represents recurrent kernel matrices of the gates; and \( b \) represents corresponding bias terms. Moreover, \( f, i, o, c, x, h \) and \( c \) represent the forget gate, input gate, output gate, vector for cell updates, input, hidden state, and cell state, respectively. Subscript \( t \) indicates the time step. For hLSTM, the recurrent layer in equation 1 is modified as follows:

**Feedforward DNN Model**
Our baseline feed forward artificial neural network—commonly called DNN—used in this project consists of three fully connected layers, a dropout layer, and an output layer. The fully connected layers have 256, 128, and 64 neurons, respectively, and they use rectified linear unit (ReLU) as the activation function. The dropout layer has a probabilistic dropout rate of 0.5. Sigmoid function was used as activation at the output layer. The loss function used was binary cross-entropy, and the optimization algorithm used was Adam. The baseline DNN model and the pruned DNN model (pDNN) were all trained for 20 epochs, using a batch size of 8. The input into the DNN model has the same feature set as LSTM model but does not consider time series information. The values were calculated by averaging nonmissing values across time steps.

**Neural Network Pruning**
All neural network prunings were conducted at the channel level, which means a neuron and all its inputs and outputs were removed from the model if the neuron is pruned. Keras surgeon library in python was used for pruning. In each layer, neurons were pruned if their mean weight across all inputs from the previous layer were below the set quantiles (ie, 25% and 50% in this study). The original model was trained for 1 epoch before pruning and was trained for another 19 epochs after pruning.

**Neural Network Quantization**
Quantization was applied on the DNN model post training. Parameters, including weights and activation, originally stored in a 32-bit floating point format were converted to 8 bits using TensorFlow Lite. A uniform quantization strategy was used, as previously described [11]. Considering the range of float point values in the model to be \( (F_{\text{min}}, F_{\text{max}}) \), all the floating-point values were quantized into the range \( (0, 255) \) as 8 bits in a uniform manner, where \( F_{\text{min}} \) corresponds to 0 and \( F_{\text{max}} \) corresponds to 255.

The quantization process is

where \( x \) is the floating-point variable, \( x_q \) is the quantized variable, and

**Results**

**Recurrent Artificial Neural Network Condensation: hLSTM and Pruned LSTM**
Recurrent artificial neural networks (or simply, RNNs) are a group of machine learning models widely used in clinical settings that take sequential or time series information as the
input. However, training of RNNs and running inference from RNNs are relatively computationally intensive. In order to enable the machine learning algorithms to be used on devices with limited computational power, such as those in high-income countries and on mobile devices, we used three strategies to reduce the storage size of the model and to increase the speed of training and inference (Figure 2).

We built a baseline RNN using two layers of LSTM neurons to predict ICU mortality rates using MIMIC-III dataset [15]. After training, the baseline RNN model achieved a decent performance of AUROC of 0.85 (Table 2).

The first strategy was to modify the LSTM cell to increase the representation power of each layer. We modified the original neural network structure and added an additional hidden layer into the original LSTM class, wherein one additional layer called “hidden kernel” was inserted between the input kernel and the recurrent kernel (see equation 4). By using this strategy, we replaced the old 2-layer LSTM with only one layer of hLSTM, such that we simplified the overall structure by trying to embed the same quantity of information in this single “condensate” layer.

Figure 2. Neural network condensation methods. (A) Hidden-layer long short-term memory (LSTM). Instead of single fix layer nonlinearity for gate control of LSTM, multiple layer neural network with ReLu as activation were used to enhance the gate controls. In this way, fewer layers of LSTMs were needed to build a model with similar performance. (B) A large portion of parameters in artificial neural networks are redundant. We pruned 50% of the channels (neurons) with the lowest weights in each layer to reduce size and complexity of the neural network. (C) Most artificial neuron network implementation in research settings uses 32- or 64-bit floating points for model parameters. We quantized the parameters to 8 bits after training to reduce sizes of the models. DNN: dense neural network.

Table 2. Recurrent neural network condensation.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters, n</th>
<th>File size (kb)</th>
<th>Inference (seconds per sample)</th>
<th>Training time (seconds; 20 epochs)</th>
<th>Test AUROCa (last epoch)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline LSTMb</td>
<td>8081</td>
<td>129</td>
<td>523</td>
<td>4890</td>
<td>0.836</td>
</tr>
<tr>
<td>Pruned LSTM</td>
<td>3273</td>
<td>73</td>
<td>318</td>
<td>4990</td>
<td>0.853</td>
</tr>
<tr>
<td>Hidden-layer LSTM</td>
<td>6993</td>
<td>111</td>
<td>254</td>
<td>3000</td>
<td>0.860</td>
</tr>
</tbody>
</table>

a AUROC: area under the receiver operating curve.
b LSTM: long short-term memory.
Both the baseline model and the hLSTM model with only one layer of hLSTM are trained under the same settings. The comparison of AUROC and accuracy is shown in Figure 3. The number of parameters for these two models are listed in Table 2. This simplified model with a single layer of hLSTM beats the baseline model 2-fold in training speed, achieving a 32% reduction in parameter numbers while simultaneously maintaining a higher AUROC at the same time.

Figure 3. Accuracy, model size, and inference speed of feedforward recurrent and neural networks (RNNs) after different types of condensation. (A) Area under the receiver operating characteristic curve (AUROC) of various models. (B) Various model sizes in memory. (C) Inference speed of various models. Models included the RNN baseline model with two layers of long-term short memory (LSTM), pruned LSTM (pLSTM) model, and one hidden layer inserted in LSTM (hLSTM); deep neural network (DNN) baseline model; pruned DNN (pDNN) model; quantized DNN (qDNN) model.

Another method to condense RNN models is pruning, in which some unessential neurons of the RNN model are removed to minimize model size and increase speed. About 50% of LSTM neurons with lowest weights in each hidden layer were pruned after the first epoch of training. The pruned LSTM only has half of the number parameters of the original LSTM, but it achieves a similar level of accuracy, yielding an AUROC of 0.85 (Figure 4). The inference speed of pruned LSTM also doubled compared with the original LSTM (Table 2).
Feedforward Neural Network Condensation: Pruning and Quantization

Feedforward neural network, or commonly called DNN if it has multiple hidden layers, is another widely used form of machine learning in clinical settings. We trained DNN with 3 hidden layers, consisting of 256, 128, and 64 neurons in each layer, to enable ICU mortality prediction. The baseline DNN achieved an AUROC of 0.82, using patient data collected within the first 48 hours after admission. We explored two methods to condense the size of the DNN. The first method, called pruning, used the pruning strategy as in RNN; for this purpose, 50% of the channels were pruned after the first epoch of training, the prediction accuracy of the pDNN maintained at the same level as the original DNN, and the inference speed doubled (Table 3). The second strategy involved quantization, which refers to the process of reducing the number of bits that represent a number. In the context of this project, the predominant numerical format used was a 32-bit floating point. We used an after-training-quantization strategy to represent the parameters of the DNN model using 8-bit integers (ie, quantized DNN or qDNN). This method reduced storage size of the DNN model by 5 times without incurring significant loss in accuracy (Table 3). We also compared the overall performances of DNN condensation with those of RNN, as shown in Figure 3.

Table 3. Feedforward neural network condensation.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters, n</th>
<th>File size (kb)</th>
<th>Inference (seconds per sample)</th>
<th>Training time (seconds; 20 epochs)</th>
<th>Test AUROC(^{a}) (last epoch)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline DNN(^{b})</td>
<td>60,929</td>
<td>767</td>
<td>20</td>
<td>3300</td>
<td>0.82</td>
</tr>
<tr>
<td>Pruned DNN</td>
<td>27,312</td>
<td>315</td>
<td>10</td>
<td>3310</td>
<td>0.81</td>
</tr>
<tr>
<td>Quantized DNN</td>
<td>60,929</td>
<td>64</td>
<td>15</td>
<td>N/A(^{c})</td>
<td>0.82</td>
</tr>
</tbody>
</table>

\(^{a}\)AUROC: area under the receiver operating curve.

\(^{b}\)DNN:

\(^{c}\)N/A: not applicable.

Discussion

In this study, we were able to use data from the MIMIC-III database [15] to train in-hospital mortality neural network models with high accuracy and conduct model condensation with different methods to gain efficiency (eg, memory size reduction and increased speed) without compromising accuracy. We implemented different neural network architectures for both RNNs and dense neural networks; thus, our methods can add value in both settings. We pioneered RNN pruning with clinical implementation and our condensation treatments aiming at higher efficiency can be extended to other medical applications using similar data, and probably to nonmedical applications as well. In addition, in medical settings, model calibration is conducted after initial model training. Calibration can be
conducted using various training schemes and early stopping strategies. The model condensation method proposed in this study significantly reduces the number of parameters and will help make model calibration easier. The major limitation of the neural network condensation method is that although our proposed method significantly reduces the sizes of different models and their computational costs in training, the final model sizes after condensation are still proportional to the original model sizes. Therefore, if further model size reduction is warranted, a combination of better model design and neural network condensation will be required.

Acknowledgments

We thank the authors of the published article “Multitask Learning and Benchmarking with Clinical Time Series Data” for providing the basic code for data preprocessing from Medical Information Mart for Intensive Care-III via GitHub. We thank Professor David Sontag (Massachusetts Institute of Technology, Cambridge, MA) for their helpful discussion.

Conflicts of Interest

None declared.

References

Abbreviations

**AUROC:** area under the receiver operating curve  
**DNN:** deep neural network  
**hLSTM:** hidden-layer long short-term memory  
**ICU:** intensive care unit  
**LSTM:** long short-term memory  
**MIMIC-III:** Medical Information Mart for Intensive Care-III  
**pDNN:** pruned deep neural network  
**qDNN:** quantized deep neural network  
**ReLU:** rectified linear unit  
**RNN:** recurrent neural network
Transitioning From In-Person to Remote Clinical Research on Depression and Traumatic Brain Injury During the COVID-19 Pandemic: Study Modifications and Preliminary Feasibility From a Randomized Controlled Pilot Study

Lauren B Fisher1,2, PhD; Sylvie Tuchman1, BA; Andrew J Curreri3, MA; Maggie Markgraf1, BS; Maren B Nyer1,2, PhD; Paolo Cassano1,2, MD, PhD; Grant L Iverson4,5,6, PhD; Maurizio Fava1,2, MD; Ross D Zafonte4,5,6,7, DO; Paola Pedrelli1,2, PhD

1Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States
2Department of Psychiatry, Harvard Medical School, Boston, MA, United States
3Department of Psychological & Brain Sciences, Boston University, Boston, MA, United States
4Department of Physical Medicine and Rehabilitation, Harvard Medical School, Boston, MA, United States
5Department of Physical Medicine and Rehabilitation, Spaulding Rehabilitation Hospital, Boston, MA, United States
6Red Sox Foundation and Massachusetts General Hospital Home Base Program, Boston, MA, United States
7Department of Physical Medicine and Rehabilitation, Massachusetts General Hospital and Brigham and Women’s Hospital, Boston, MA, United States

Corresponding Author:
Lauren B Fisher, PhD
Department of Psychiatry
Massachusetts General Hospital
1 Bowdoin Square
6th Floor
Boston, MA, 02114
United States
Phone: 1 617 643 0877
Email: fisher.lauren@mgh.harvard.edu

Abstract

Background: Telehealth has provided many researchers, especially those conducting psychosocial research, with the tools necessary to transition from in-person to remote clinical trials during the COVID-19 pandemic. A growing body of research supports the effectiveness of telemental health for a variety of psychiatric conditions, but few studies have examined telemental health for individuals with comorbid medical diagnoses. Furthermore, little is known about the remote implementation of clinical trials examining telemental health interventions.

Objective: This paper outlines the procedural modifications used to facilitate conversion of an in-person randomized controlled trial of cognitive behavioral therapy (CBT) for depression in individuals with traumatic brain injury (TBI; CBT-TBI) to a telemental health study administered remotely.

Methods: Given the nature of remote implementation and specific challenges experienced by individuals with TBI, considerations related to treatment delivery, remote consent, data management, neuropsychological assessment, safety monitoring, and delivery of supportive material have been discussed. Feasibility, acceptability, and safety were evaluated by examining attendance and participant responses on self-report measures of treatment satisfaction and suicidal behavior.

Results: High rates of treatment attendance, assessment completion, study retention, and satisfaction with the intervention and modality were reported by participants who completed at least one telemental health CBT-TBI session.

Conclusions: Study modifications are necessary when conducting a study remotely, and special attention should be paid to comorbidities and population-specific challenges (eg, cognitive impairment). Preliminary data support the feasibility, acceptability, and safety of remotely conducting a randomized controlled trial of CBT-TBI.

Trial Registration: ClinicalTrials.gov NCT03307070; https://clinicaltrials.gov/ct2/show/NCT03307070
Introduction

The COVID-19 pandemic has had significant impacts on the conduct of clinical research. As a result of stay-at-home orders, most in-person clinical trials for nonsalvaging interventions were suddenly halted, leaving many active research participants and teams in limbo. Many researchers were forced to choose between pausing their research projects, with a threat to scientific productivity, and modifying procedures to implement remote approaches [1].

Psychosocial research (eg, investigation of psychological treatments) is well positioned to be conducted remotely. Telemental health, the application of telecommunications to provide mental health services from a distance, has grown exponentially during the COVID-19 pandemic [2] and includes the use of a wide range of technologies to deliver synchronous (eg, live videoconferencing or telephone calls) and asynchronous interventions (eg, web-based interventions completed without a clinician present) [3]. This paper primarily focuses on the use of synchronous exchanges with telephone and videoconferencing to facilitate the remote implementation of assessment and psychotherapy in the context of a clinical trial for depression that was conducted in-person prior to the pandemic. The transition to remote study implementation was supported by a growing body of research demonstrating the effectiveness of telemental health services across many populations (eg, adults, children, and older adults) and for a range of psychiatric conditions [4], including depression [5], anxiety [6], and posttraumatic stress disorder [7]. In fact, real-time telemental health (ie, videoconferencing or telephone) is as effective as face-to-face treatment in reducing depressive symptoms [3,8]. Furthermore, treatment satisfaction and therapeutic alliance are similar among patients engaged in telemental health (videoconferencing and telephone-based interventions) and in-person treatment [9]. Effective implementation of protocols for remotely assessing and managing suicide risk further supports the feasibility of conducting clinical trials that examine telemental health interventions for individuals with depression [10].

Despite the significant promise of telemental health for many individuals with depression, individuals with various comorbid medical diagnoses may experience distinct challenges that serve as barriers to effectively utilizing telemental health interventions and participating in clinical research implemented remotely. In our work with individuals who have sustained traumatic brain injury (TBI), the impacts of TBI sequelae, including cognitive difficulties (eg, impaired focus and attention, and executive dysfunction) and sensitivity to light/screens, present unique challenges to participation in telemental health. Nevertheless, preliminary evidence suggests that individuals with major depressive disorder (MDD) and complicated mild to severe TBI experience similar reductions in depressive symptoms after 16 weeks of in-person and telephone-delivered cognitive behavioral therapy (CBT), and report similarly high rates of treatment satisfaction and strong therapeutic alliance [11]. Although we have been unable to identify studies examining the use of videoconferencing for the delivery of individual psychotherapy for adults with depression and TBI, there is support for the feasibility of using videoconferencing for group CBT to improve emotion regulation after TBI [12] and problem solving–based telemental health for improving behavior and family functioning in children with TBI [13]. Furthermore, support for the feasibility of video-based telehabilitation in adults with TBI [14] suggests that telemental health interventions adapted for this population may also be feasible.

Methods

Participants

Study participants were enrolled in an ongoing randomized waitlist-controlled trial (target N=40) piloting a 12-week individual CBT for depression that was adapted for individuals with TBI. The aims of the parent trial were to evaluate the feasibility and acceptability of the intervention (primary), as well as the potential efficacy in reducing depressive severity (secondary). As of August 27, 2021, the ongoing RCT enrolled (consented) a total of 33 participants, of which 18 participants were enrolled in-person and 15 participants were enrolled remotely. The clinical trial began in-person recruitment in April 2019 and transitioned to remote procedures on March 16, 2020, which remains ongoing at the time of writing. Of the 33 enrolled participants, 8 completed all study visits in-person, 6 completed a combination of in-person and telemental health sessions, 2 were enrolled in-person but completed all CBT-TBI sessions remotely, 7 completed all study visits remotely, 3 were deemed ineligible at the screening session (1 in-person and 2 remote), 2 discontinued (during remote CBT-TBI), 1 was lost to follow-up, and 3 determined they did not have time to participate (immediately after remote enrollment). One participant remains active at the time of writing. The feasibility and safety analyses presented below include participants who completed one or more CBT-TBI sessions remotely. Acceptability data include participants who completed one or more CBT-TBI sessions.
remotely and completed the study (defined as attending all 12 intervention sessions; n=12), as well as 1 participant who terminated CBT-TBI early but completed end-of-study assessments. Participants in this subsample of the ongoing RCT (n=18) were between the ages of 21 and 69 years (mean 39.2 years, SD 16.3 years), and the majority were white (n=14, 78%), non-Hispanic or Latino (n=15, 83%), and highly educated (n=10, 56% with at least 4 years of college). Just over half of the sample (n=10, 56%) were women, while less than half the sample (n=8, 44%) were married or in a relationship.

Procedures

All study procedures, including pandemic-related modifications, were approved by the Massachusetts General Hospital Institutional Review Board. Adults with clinically significant depressive symptoms and a history of moderate-to-severe TBI were included in the study (see Multimedia Appendix 1 for the full study criteria [15-18]).

The initial screening visit included informed consent, diagnostic and symptom evaluation with a study clinician, and a neuropsychological battery that was completed in-person prior to March 2020 and remotely since April 2020. Participants also completed a series of baseline self-report measures of mood; suicidality; and cognitive, social, and emotional functioning, using a secure web-based platform (REDCap). Eligible participants were then randomized to 12 weeks of a newly developed manualized cognitive behavioral treatment for depression adapted for individuals with TBI (CBT-TBI) or a 12-week waitlist. The intervention included psychoeducation, behavioral activation, goal setting, cognitive restructuring, and relapse prevention. CBT-TBI was adapted for individuals with TBI by incorporating the following strategies: repetition, patient workbook with session summaries and forced choice worksheets, modified thought records, therapeutic use of neuropsychological testing results, individually tailored text messages/between-session reminders, and daily use of an activity monitoring device (Fitbit Charge 3). Individuals in both conditions completed bimonthly phone assessments of depressive symptoms with an independent evaluator. Weekly 50 to 60-minute individual CBT-TBI sessions were delivered by a master’s or doctoral-level clinician in-person until March 16, 2020, and via Zoom videoconferencing (or telephone, when needed) thereafter. At the end of 12 weeks, all participants completed a postassessment, which included clinician-rated and self-rated measures, as well as repeat neuropsychological testing. Individuals randomized to the waitlist condition could receive CBT-TBI upon completion of 12 weeks of assessment. All study procedures are outlined in Figure 1, and a detailed breakdown of CBT-TBI visits is shown in Figure 2.

Figure 1. Flowchart of study procedures for eligible participants. *All participants randomized to waitlist can complete the intervention following the final assessment. CBT-TBI: cognitive behavioral therapy for depression in individuals with traumatic brain injury.
**Figure 2.** Flowchart of study procedures for treatment phase. BDI-II: Beck Depression Inventory-II; CBT-TBI: cognitive behavioral therapy for depression in individuals with traumatic brain injury.

**Measures**

**Acceptability**

The 12-item self-rated Satisfaction with Therapy and Therapist Scale-Revised (STTS-R) [19] was used to assess satisfaction in 2 domains of treatment. Current analyses included the satisfaction with therapy subscale scores, which range from 6 to 30, with higher scores indicating greater satisfaction. In March 2020, 5 questions were composed by the study team to gather feedback on remote CBT-TBI visits, including satisfaction with remote CBT-TBI sessions on a 5-point Likert scale (1, very satisfied to 5, very dissatisfied). Participants were also asked to share what they liked and did not like about virtual treatment. Participants who completed some CBT-TBI sessions in-person and some remotely were asked to indicate the degree of their preference for one modality over the other on a 5-point Likert scale (1, strongly preferred telemental health sessions to 5, strongly preferred in-person sessions). Finally, participants were asked to select the modality they would choose if given the option for treatment after the pandemic (eg, in-person, over the telephone, videoconferencing, and combination of in-person and virtual).

**Safety**

Suicidal ideation was monitored weekly during CBT-TBI with the suicide item from the Beck Depression Inventory-II (BDI-II) [20], a 21-item self-report scale designed to measure the presence and severity of depressive symptoms. The BDI-II suicide item is associated with the risk of repeat suicide attempts and death by suicide and is recommended as a screener for suicide risk in routine clinical care [21]. Adverse events were also assessed during bimonthly phone assessments.

**Study Modifications With Transition to Remote Implementation of Research**

Several protocol modifications (Table 1) were instituted after all clinical trials for nonlifesaving interventions were halted in our institution due to the pandemic. Modifications aimed to facilitate feasibility and adherence to the original procedures as much as possible.
<table>
<thead>
<tr>
<th>Protocol element</th>
<th>In-person implementation</th>
<th>Remote implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment modality</td>
<td>Individual face-to-face sessions in the office</td>
<td>Individual sessions via secure videoconferencing</td>
</tr>
<tr>
<td>Consent</td>
<td>Clinicians and participants review paper consent and sign in the office</td>
<td>Clinicians and participants utilize teleconsent with the REDCap eConsent template during videoconference</td>
</tr>
</tbody>
</table>
| Data management                  | • Participants complete questionnaires directly on REDCap using an in-office computer (preferred method)  
  • REDCap links are emailed to participants who are unable to complete questionnaires during office visit  
  • Paper copies are mailed with self-addressed return envelope for participants unable to complete electronically  
  • Clinicians complete pencil and paper assessments (requires data entry) | • Via text or email, participants are sent a REDCap link to complete questionnaires independently (preferred method)  
  • Paper copies are mailed with self-addressed return envelope for participants unable to complete electronically  
  • Clinicians enter clinical data directly into REDCap |
| Neuropsychological assessment    | • Administered by study staff in the office (traditional measures):  
  • TOPF\textsuperscript{a} \[22\]  
  • WAIS-IV\textsuperscript{b} Coding, Digit Span and Similarities \[23\]  
  • D-KEFS\textsuperscript{c} Color Word and Trails \[24\]  
  • CVLT-II\textsuperscript{d} \[25\]  
  • Administered on an iPad in the office:  
  • NIH\textsuperscript{e} Toolbox Cognition Battery \[26\] | • Administered by study staff via videoconferencing (traditional measures):  
  • TOPF \[22\]  
  • WAIS-IV Digit Span and Similarities  
  • WMS-IV\textsuperscript{f} Logical Memory I and II \[27\]  
  • Computerized battery completed by participants at home  
  • CNS Vital Signs \[28\] |
| Suicide risk monitoring          | Clinicians review the paper copy responses to the BDI-II suicide item at the start of the CBT-TBI visit with participants in the room | The study coordinator reviews REDCap responses to the BDI-II suicide item at the start of the CBT-TBI visit and alerts clinicians to scores of 2 or higher |
| Preparation for CBT-TBI visits   | Routine scheduling; the study coordinator answers questions from participants             | • A “Welcome Letter” is sent to establish expectations:  
  • Ensure security (eg, close other applications while Zoom is open)  
  • Ensure privacy (eg, conduct sessions in a private room with the door closed, use headphones and/or a noise blocker)  
  • Provides tips for limiting distractions (eg, silence cell phone, avoid eating, ensure device is fully charged, and let others in the home/space know you are unavailable)  
  • Consider the feasibility of your device (eg, a computer allows for typing notes in electronic handouts and hardwired ethernet connections can be more reliable than Wi-Fi) |
| Delivery and setup of wearable technology | The study coordinator sets up Fitbit with participants on the day of the first in-office CBT-TBI session | The study coordinator mails the device to participants and guides them through device setup via videoconferencing |
| CBT-TBI delivery modification    | Provide handouts in the session that are added to the CBT-TBI Workbook every week         | • Mail the CBT-TBI Workbook with handouts and worksheets prior to the start of treatment  
  • Minimize reliance on screens (eg, turn away from the computer and turn off video)  
  • Tailor delivery to individual needs/preferences and be flexible (eg, utilize “screen share” and provide electronic handouts) |

\textsuperscript{a}TOPF: Test of Premorbid Functioning.  
\textsuperscript{b}WAIS-IV: Wechsler Adult Intelligence Scale–Fourth Edition.  
\textsuperscript{c}D-KEFS: Delis-Kaplan Executive Function System.  
\textsuperscript{d}CVLT-II: California Verbal Learning Test–Second Edition.  
\textsuperscript{e}NIH: National Institutes of Health.  
\textsuperscript{f}WMS-IV: Wechsler Memory Scale–Fourth Edition.
Videoconferencing Platform

All procedures (except for neuropsychological testing, discussed below) that were previously conducted face-to-face were completed remotely using videoconferencing. Secure Zoom videoconferencing (Zoom Enterprise) was adopted after working with our institution’s Research Information Security Office to optimize privacy and security settings, including enabling the waiting room, locking meetings once sessions begin, and generating meeting IDs with a password. Individuals who were unfamiliar with the videoconferencing platform received step-by-step instructions for Zoom account set-up, and could participate in a “trial run” and orientation to the platform with the study coordinator.

Remote Consent

Since the study transitioned to remote implementation, the informed consent process was embedded into a live telehealth session, also referred to as teleconsent [29]. An institution-specific REDCap electronic informed consent template was utilized. The study clinician met with the participant over Zoom to review the consent form and instruct the participant to digitally sign consent. A signed copy was then securely emailed directly to the participant from REDCap. Participants were given the option to receive a mailed paper copy and/or a brief summary of key study information to make the process less overwhelming. Overall, teleconsent provides a feasible alternative to in-person paper consent and facilitates research continuity when face-to-face interactions are not possible [30].

Data Management

Prior to the pandemic, study participants were given the option of completing self-report questionnaires directly in REDCap using a computer or tablet, either in our office or at home. Individuals participating remotely were provided an electronic link to complete questionnaires at home directly on REDCap. Given that it can be cognitively taxing to sit at a screen for an extended duration of time, participants were encouraged to complete a few questionnaires at a time. Participants were offered the option of being mailed paper questionnaires with a self-addressed envelope that was returned to the research team.

Neuropsychological Testing

The original in-person battery included a series of traditional paper and pencil neuropsychological measures and the iPad-administered NIH Toolbox Cognition battery [26]. Following a review of the available teleneuropsychology literature [31] and guidelines for remote assessments [32], it was determined that certain subtests from the original battery could be administered via videoconferencing, likely without significant impact on reliability and validity, although some tests could not (Table 1). CNS Vital Signs [28], a brief computerized neuropsychological test battery, replaced the NIH Toolbox Cognition battery and was administered remotely in accordance with guidelines to maximize validity. Participants were instructed to watch a preparatory video that emphasizes the importance of creating an optimal standardized testing environment (eg, limit distractions and interruptions, and set aside sufficient time to complete), which in turn maximizes the reliability of test results. CNS Vital Signs reports include a validity indicator for each subtest, allowing the clinician to follow-up with the participant about possibly invalid results.

Suicide Risk Assessment

On the day of CBT-TBI sessions, the study coordinator emailed or texted the REDCap link for the BDI-II [20] to the participants, instructed them to complete the measure prior to meeting with the study therapist, and reviewed their responses to the BDI-II suicide item in real-time. If the participant did not complete the measure, the study therapist was notified to remind the participant to do so and to review the suicide item response before starting the session. In the event that the participant had a score of 2 or higher, the therapist was immediately alerted to conduct a detailed suicide risk assessment in the session and to determine the need for a higher level of care, which could involve voluntary or involuntary hospitalization and/or contacting the individual’s previously identified emergency contact. The participant’s physical location was identified after signing consent and was confirmed before starting every CBT-TBI session. All efforts were made for the study therapist to remain connected to the participant (on Zoom or telephone) until emergency personnel arrived at their location.

Preparing for CBT-TBI Telehealth Visits

Given that individuals with TBI can be sensitive to changes in routine due to deficits in mental flexibility and problem solving [33], it is encouraged that telemental health visits mirror in-person CBT-TBI visits as much as possible. A predictable environment that parallels the in-person setting (ie, consistent office space/background) may be beneficial [34]. In order to compensate for a patient’s reduced ability to read the therapist’s nonverbal cues over video, clinicians configured their camera to ensure that the patient could see as much of their body language as was feasible [34]. Clear and consistent expectations about virtual visits were directly communicated in a “Welcome Letter” that emphasized how best to ensure security and privacy, and provided tips for limiting distractions during the session (Table 1). Participants who had more than one internet and video-enabled device were encouraged to consider which device would best suit their needs based on factors such as strength of internet connection and device portability. Additionally, the active hands-on nature of the study intervention, which uses worksheets and encourages notetaking, warranted an appropriate workspace, such as sitting at a desk or table.

Delivery and Setup of Wearable Technology

Participants were mailed a Fitbit Charge 3 activity tracker prior to starting the intervention. The study coordinator scheduled individual videoconferencing meetings with participants prior to initiating therapy. To ensure the participant understood how to use the device, the researcher scheduled a “trial run” on REDCap to support the participant in becoming comfortable with the platform and to verify that the participant had access to a reliable internet connection. The researcher reviewed step-by-step instructions for setting up an account, and emphasized the importance of setting aside sufficient time to use the device. Given that it can be cognitively taxing to sit at a screen for an extended duration of time, participants were encouraged to complete a few questionnaires at a time. Participants were offered the option of being mailed paper questionnaires with a self-addressed envelope that was returned to the research team.

Delivery and Setup of Wearable Technology

Participants were mailed a Fitbit Charge 3 activity tracker prior to starting the intervention. The study coordinator scheduled individual videoconferencing meetings with participants prior to initiating therapy. To ensure the participant understood how to use the device, the researcher scheduled a “trial run” on REDCap to support the participant in becoming comfortable with the platform and to verify that the participant had access to a reliable internet connection. The researcher reviewed step-by-step instructions for setting up an account, and emphasized the importance of setting aside sufficient time to use the device. Given that it can be cognitively taxing to sit at a screen for an extended duration of time, participants were encouraged to complete a few questionnaires at a time. Participants were offered the option of being mailed paper questionnaires with a self-addressed envelope that was returned to the research team.

Delivery and Setup of Wearable Technology

Participants were mailed a Fitbit Charge 3 activity tracker prior to starting the intervention. The study coordinator scheduled individual videoconferencing meetings with participants prior to initiating therapy. To ensure the participant understood how to use the device, the researcher scheduled a “trial run” on REDCap to support the participant in becoming comfortable with the platform and to verify that the participant had access to a reliable internet connection. The researcher reviewed step-by-step instructions for setting up an account, and emphasized the importance of setting aside sufficient time to use the device. Given that it can be cognitively taxing to sit at a screen for an extended duration of time, participants were encouraged to complete a few questionnaires at a time. Participants were offered the option of being mailed paper questionnaires with a self-addressed envelope that was returned to the research team.

Delivery and Setup of Wearable Technology

Participants were mailed a Fitbit Charge 3 activity tracker prior to starting the intervention. The study coordinator scheduled individual videoconferencing meetings with participants prior to initiating therapy. To ensure the participant understood how to use the device, the researcher scheduled a “trial run” on REDCap to support the participant in becoming comfortable with the platform and to verify that the participant had access to a reliable internet connection. The researcher reviewed step-by-step instructions for setting up an account, and emphasized the importance of setting aside sufficient time to use the device. Given that it can be cognitively taxing to sit at a screen for an extended duration of time, participants were encouraged to complete a few questionnaires at a time. Participants were offered the option of being mailed paper questionnaires with a self-addressed envelope that was returned to the research team.
to the first CBT-TBI session where they provided instruction on setup and use, and provided a handout reiterating this information.

Minimize Reliance on Screens

For some individuals with TBI, screen time can exacerbate symptoms, including headaches due to photosensitivity [35]. To reduce the degree to which participants engaged with material in electronic format, all participants were mailed a physical copy of the CBT-TBI Workbook, which contained copies of weekly agendas, handouts, and worksheets. Participants who reported considerable difficulty with photosensitivity were encouraged to turn away from the computer or turn off video.

Importance of Tailoring Delivery and Being Flexible

Given the importance of clear direct communication for individuals experiencing cognitive difficulties, the therapist spoke with the participants in the first session to understand their comfort with technology and preferences for ways of engaging in the collaborative treatment (eg, physical vs electronic worksheets, therapist vs participant typing responses into worksheets, and use of the videoconferencing screen share feature to provide visual cues and allow the therapist to model skill use). Consistent with procedures utilized during in-person delivery of the intervention, participants continued to receive between-session reminders via text, email, or phone call to carry out collaboratively identified activities or goals (eg, behavioral activation).

Troubleshooting Challenges With Technology

It is inevitable that technological challenges will arise both prior to and during virtual sessions. It is important that clinicians do not get visibly frustrated in the face of technological difficulties, as the patient may interpret this as the clinician being upset with them [34]. Additionally, shared insecurities over technology may, in fact, aid therapeutic alliance [36]. However, it is worth recognizing that technological issues can be disruptive, and clinicians may want to identify a cutoff point at which they switch from videoconferencing to a telephone session. In the first CBT-TBI session, the clinician and participant develop an individualized plan for navigating potential technological difficulties, such as losing a connection mid-session.

Data Analyses

To assess the preliminary feasibility and acceptability of remote study implementation, descriptive statistics were used to report the number of CBT-TBI intervention sessions attended, number of assessment sessions attended, number of participants who completed the study, and rate of satisfaction with treatment (STTS-R and supplemental questions). Study retention (number of participants/number randomized) was also calculated. Finally, the number of elevated responses to the BDI-II suicide item (≥2) was described to demonstrate the maintenance of safety protocols.

Results

Feasibility

At the time of the transition to remote procedures (March 2020), there were 9 active study participants, including 3 participants on the waitlist (ie, had not started CBT-TBI) and 6 participants who were mid-treatment. The 6 participants who started CBT-TBI in-person prior to March 2020 were at different points in treatment at the time of the transition (weeks 3, 7, 9, 10 11, and 12), and all went on to complete the remainder of their 12 weeks of CBT-TBI remotely using telemental health sessions. Two out of three waitlist participants who enrolled in the study before March 2020 with the expectation of attending CBT-TBI sessions in-person completed the entire 12 weeks of CBT-TBI via videoconferencing. One participant discontinued participation after 4 CBT-TBI sessions due to a demanding work schedule but completed postassessments. Finally, of all randomized participants who enrolled in the study remotely (n=9), 7 (78%) completed the study, 1 (11%) was withdrawn due to worsening depression, and 1 (11%) remains active in CBT-TBI. The study retention rate prior to March 2020 was 100% (8 CBT-TBI completers), and completion from March through January 2021 was about 93% (13 out of 14 possible randomized CBT-TBI completers). Approximately 91% of clinician-rated assessments (102 out of 112 possible assessments) were completed throughout the period in which the study had been conducted remotely.

Acceptability

Satisfaction with therapy (STTS-R therapy subscale) was high (mean 27.1, SD 2.8) among participants who completed at least one session of CBT-TBI remotely (n=16). Overall satisfaction with telemental health sessions (videoconferencing, telephone, or a combination of both) was high (n=14); 9 participants (64%) reported being “very satisfied,” 4 participants (29%) were “satisfied,” and 1 participant (7%) reported being “neither satisfied nor dissatisfied.” If given a choice of modality in the future (n=14), 3 participants (21%) indicated that they would choose in-person treatment, 4 participants (29%) indicated that they would choose telemental health treatment (videoconferencing or telephone), and 7 participants (50%) indicated that they would choose a combination of in-person and telemental health treatment. Among the participants who completed a combination of in-person and telemental health treatment and provided feedback (n=4), there was no clear pattern in preferred modality, as they reported strongly preferring telemental health treatment (n=1, 25%), somewhat preferring telemental health treatment (n=1, 25%), strongly preferring in-person treatment (n=1, 25%), and no indication of preference (n=1, 25%).

Qualitative feedback highlighted that all 14 study completers noted at least one benefit of telemental health sessions, including ease of conducting sessions from home and not having to travel for appointments. Conversely, technological challenges, reduced focus, limited privacy, and difficulty feeling connected with the therapist were noted as factors that participants disliked about telemental health sessions. Five participants reported that there was nothing they disliked about telemental health treatment.
Safety
Since March 2020, 1 participant had a score of 3 on the BDI-II [20] suicide item in 2 consecutive weeks. Per protocol, the study coordinator alerted the CBT-TBI therapist and the principal investigator immediately after having identified the safety concern, and the therapist started the CBT-TBI session with a thorough assessment of suicide risk. After several weeks of worsening depression and increasing suicidal ideation, the participant was eventually referred for a higher level of care (partial hospitalization program), was withdrawn from the study, and was ultimately hospitalized voluntarily for worsening of symptoms. No serious adverse events were reported throughout the duration of remote procedures.

Discussion
Principal Findings
Using several procedural modifications described in this paper, an in-person RCT of CBT for depression after TBI was converted to remote implementation and demonstrated preliminary evidence of feasibility, acceptability, and safety. Specific modifications to study implementation and the treatment protocol are outlined in Table 1. Given the range of symptoms and deficits that can arise after TBI (eg, photosensitivity and impaired attentional capacity), all modifications were made with consideration of their potential impact on participants and to enhance feasibility.

Preliminary data supported the feasibility, acceptability, and safety of conducting an RCT of CBT for depression among individuals with TBI exclusively utilizing remote procedures. Specifically, preliminary results demonstrated a high rate of completion for clinician-rated assessments (102/112, 91%) and high study retention (13/14, 93%). Procedures that were designed to monitor safety were effective in identifying individuals at high risk for suicide, triggering clinician suicide risk assessments via videoconferencing. Feedback from participants suggested a high degree of satisfaction with the CBT-TBI treatment and telemental health modality, providing initial evidence of the acceptability of the remotely delivered study intervention. The findings are consistent with the results of previous studies that have examined telephone-delivered cognitive behavioral interventions among individuals with TBI [11,37].

Feedback from our small sample highlighted a range of preferences when participants were asked to consider their ideal treatment modality (in-person treatment, telemental health treatment, or a hybrid model), which has significant implications for study participation and potentially for treatment outcomes. Research has demonstrated better treatment outcomes among individuals whose preferences about psychological treatment (eg, appointment time, venue, and treatment type) are accommodated compared with individuals whose preferences are not met [38]. Previous research among depressed individuals with TBI utilized choice-stratified randomization, in which participants could assert a preference for CBT that was delivered in-person or over the telephone prior to randomization, in order to enhance ecological validity [11]. Qualitative feedback from our study suggested that participants may appreciate a mix of in-person and telemental health visits, which is consistent with evidence for the high feasibility and acceptability of “blended” models of delivery (combination of face-to-face and web-based sessions) of CBT for depression [39]. Although the efficacy of our study intervention is unknown at this time, tailoring the intervention modality according to preferences may lead to greater attendance at treatment sessions and engagement in treatment.

Limitations
It is important to acknowledge several limitations. Although several steps were taken to optimize the testing environment, neuropsychological assessment is ideally suited to in-person administration. Challenges with technology and suboptimal conditions in the participant’s environment have the potential to impact engagement and data collection. Behavioral observations can be restricted by videoconferencing, and rapport can sometimes be limited without in-person interactions, which may impact participant responses or commitment to participation, especially prior to randomization. It is also important to note that our sample was heavily comprised of individuals who received specialized acute inpatient rehabilitation (n=11, 61%) and specialized outpatient treatment (n=13, 72%) for their TBIs in a single academic medical center in the Northeast. Individuals who receive inpatient rehabilitation represent 7% of all persons hospitalized with moderate-to-severe TBI, are less likely to be a member of a racial/ethnic minority group, and are more likely to have health insurance compared with individuals who are hospitalized and do not receive inpatient rehabilitation after moderate-to-severe TBI [40]. Thus, our sample may not be representative of all individuals with moderate-to-severe TBI in the United States.

Conclusion
Remote study participation has been a feasible alternative when in-person research was halted during the COVID-19 pandemic. Strategic procedural modifications outlined in this paper have been instrumental to the continued feasibility of recruitment and retention of individuals with depression and TBI in the context of our ongoing RCT. Furthermore, telemental health offers significant advantages in eliminating common barriers to study participation, including transportation, time needed to travel to appointments, distance to the hospital, limited mobility, and inclement weather. Conversely, some individuals may struggle to secure private space for their sessions, and a visit to a traditional office space may be preferred. Further, many individuals do not have internet access and a camera-enabled device for videoconferencing. Some individuals may find it easier and less intimidating to be vulnerable about the challenges they face over a computer screen rather than in-person [41], while others may have difficulty connecting with a therapist through a screen. For some individuals, the flexibility of utilizing both types of modalities within the course of treatment may be an ideal balance; thus, future research designs should consider the role of patient preference. For individuals with TBI who frequently struggle with physical, cognitive, and emotional impairments, flexibly tailored treatments that utilize telemental health and in-person modalities are likely to be important in both research and clinical settings. Future research should directly compare the feasibility and efficacy of CBT delivered...
via telemental health, in-person, and hybrid models for individuals with TBI, as well as the validity and reliability of remote neuropsychological assessments and strategies to facilitate remote engagement. The COVID-19 pandemic abruptly presented researchers with unique challenges that have required flexibility and innovation. The advantages presented by the ability to conduct clinical research using remote methods are likely to persist long after the pandemic ends.

Acknowledgments
The research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under award number K23HD087464. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest
PC co-founded and consults for Niraxx Light Therapeutics Inc. He has no conflicts of interest related to this publication. GLI serves as a scientific advisor for NanoDx, Sway Operations, LLC, and Highmark, Inc. He has a clinical and consulting practice in forensic neuropsychology, including expert testimony, involving individuals who have sustained mild TBIs (including athletes). He has received research funding from several test publishing companies, including ImPACT Applications Inc, CNS Vital Signs, and Psychological Assessment Resources (PAR, Inc). He has received research funding as a principal investigator from the National Football League, and salary support as a collaborator from the Harvard Integrated Program to Protect and Improve the Health of National Football League Players Association Members. He acknowledges unrestricted philanthropic support from ImPACT Applications, Inc, the Mooney-Reed Charitable Foundation, the National Rugby League, and the Spaulding Research Institute. These entities were not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication. RDZ received royalties from Springer/Demos publishing for serving as co-editor of the text Brain Injury Medicine. RDZ serves on the Scientific Advisory Board of Myomo and Onecate.ai. He has no conflicts of interest related to this publication. The other authors have no conflicts relevant to this publication.

Multimedia Appendix 1
Inclusion/exclusion criteria for the randomized controlled trial.
[DOCX File, 30 KB - formative_v5i12e28734_app1.docx ]

Multimedia Appendix 2
CONSORT checklist.
[DOC File, 219 KB - formative_v5i12e28734_app2.doc ]

References


Abbreviations

BDI-II: Beck Depression Inventory-II
CBT: cognitive behavioral therapy
CBT-TBI: cognitive behavioral therapy for depression in individuals with traumatic brain injury
MDD: major depressive disorder
NIH: National Institutes of Health
RCT: randomized controlled trial
STTS-R: Satisfaction with Therapy and Therapist Scale-Revised
TBI: traumatic brain injury

©Lauren B Fisher, Sylvie Tuchman, Andrew J Curreri, Maggie Markgraf, Maren B Nyer, Paolo Cassano, Grant L Iverson, Maurizio Fava, Ross D Zafonte, Paola Pedrelli. Originally published in JMIR Formative Research (https://formative.jmir.org), 01.12.2021. 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.
Depression, Anxiety, and Daily Activity Among Adolescents Before and During the COVID-19 Pandemic: Cross-sectional Survey Study

Anna Jolliff¹, MS; Qianqian Zhao¹, MS; Jens Eickhoff², PhD; Megan Moreno¹, MD, MSEd, MPH
Department of Pediatrics, School of Medicine and Public Health, University of Wisconsin - Madison, Madison, WI, United States

Abstract

Background: The COVID-19 pandemic has resulted in significant changes to adolescents’ daily lives and, potentially, to their mental health. The pandemic has also disproportionately affected historically marginalized and at-risk communities, including people of color, socioeconomically disadvantaged people, people identifying as female, and youth.

Objective: This study aimed to understand differences in depression and anxiety among 2 groups of adolescents in the United States before and during the COVID-19 pandemic, and to examine demographic and daily activity variables associated with depression and anxiety.

Methods: Online surveys were distributed in 2019 and 2020. Demographic questions were asked at the time of enrollment, and included participants’ age, gender, race and ethnicity, and socioeconomic status (SES). The 8-item Patient Health Questionnaire was used to assess symptoms of depression, and the 7-item Generalized Anxiety Disorder scale was used to assess symptoms of anxiety. A total of 4 pandemic-specific daily activity questions were asked only of the pandemic group. Analyses of covariance compared depression and anxiety between prepandemic and pandemic groups. Demographic and lifestyle variables were included as covariates.

Results: The sample comprised a total of 234 adolescents, with 100 participants in the prepandemic group and 134 participants in the pandemic group. Within the pandemic group, 94% (n=126) of adolescents reported being out of school due to the pandemic, and another 85.8% (n=115) and 57.1% (n=76) were prevented from extracurricular activities and exercise, respectively. Higher depression was seen in the pandemic group, with a least-squares adjusted mean of 7.62 (SD 1.36) compared to 6.28 (SD 1.42) in the prepandemic group, although the difference was not significant (P=.08). There was no significant difference in anxiety scores between the 2 groups (least-squares adjusted means 5.52, SD 1.30 vs 5.01, SD 1.36; P=.48). Within the pandemic group, lower SES was predictive of anxiety, such that those in the pandemic group of lower SES were more anxious than their higher-SES peers (least-squares adjusted means 11.17, SD 2.34 vs 8.66, SD 2.16; P=.02). Within the pandemic group, being out of work or school and not partaking in extracurricular activities or exercise due to the pandemic were not associated with higher depression or anxiety scores.

Conclusions: In this study, neither being in the pandemic group nor experiencing changes in daily activity due to the pandemic was associated with higher depression or anxiety. However, we found that adolescents from lower SES backgrounds experienced significantly more anxiety during the pandemic than their more privileged peers. Both instrumental and mental health interventions for low-income adolescents are imperative.

(JMIR Form Res 2021;5(12):e30702) doi:10.2196/30702
KEYWORDS
COVID-19; pandemic; adolescent; depression; anxiety; socioeconomic status; survey; mental health

Introduction

Background
COVID-19 resulted in significant changes to adolescents’ daily lives [1]. This pandemic led to a prolonged period of statewide school closures [2,3], social isolation, unemployment [4], millions of infections, and hundreds of thousands of deaths [5]. The pandemic also disproportionately affected historically marginalized and at-risk communities, including people of color, socioeconomically disadvantaged people, people identifying as female, and youth [6,7]. However, it is unclear whether the early days of the pandemic were associated with appreciable changes in adolescents’ mental health in the United States.

Early international research suggested that the COVID-19 pandemic was associated with elevated rates of depression and anxiety in adolescents [8]. Research performed in Spain and Italy found increases in anxiety, irritability, and restlessness among adolescents [9], while Chinese adolescents reported increases in depression and anxiety during the pandemic [8,10]. Research in Germany suggests that increases in depression, anxiety, and distress were especially pronounced among those with pre-existing mental illnesses [11]. However, much of the research on the pandemic thus far has focused on either adults or adolescents outside the United States [9]. Limited research on US adolescents during COVID-19 demonstrates increases in depression [12,13] and anxiety [13,14]. A systematic review published in 2021 [15] demonstrated that research on US adolescents was still in a nascent stage; out of 16 international studies on the psychological impacts of COVID-19 on children, only 2 focused on US participants [16,17]. This is an important gap, as the experience of American adolescents has been somewhat unique from that of adolescents in other countries; safety measures were implemented more slowly and with greater variability in the United States, and remained in effect longer than anticipated [18].

Adolescents did not all fare equally well during the COVID-19 pandemic. Existing research suggests that, during COVID-19, adolescent females scored higher in depression and anxiety than adolescent males [8,14,15,19-21]. However, because most of these studies are cross-sectional, it is not clear whether COVID-19 disproportionately affected female over male mental health, or whether the greater anxiety and depression among female adolescents simply reflects pre–COVID-19 differences [22]. Indeed, 1 longitudinal study found that although adolescent females reported greater depression than males during the pandemic, this was proportionate to their pre–COVID-19 differences in depression [20].

Existing research shows that people of lower socioeconomic status (SES) had worse mental health outcomes during COVID-19. Research shows that US adults with a lower income and less than $5000 in savings experienced greater depression during the pandemic [12]. A study of US adolescents showed that residing in a county with higher poverty levels during COVID-19 was associated with greater depression and anxiety [21]. These negative mental health consequences may be due to the disproportionately negative impact COVID-19 had on people of lower SES. This group experienced a higher incidence of COVID-19 [23] due to limited access to health information, lack of protective personal equipment, and reduced freedom to socially distance [24]. Once infected, lower SES also predicted invasive mechanical ventilation, intensive care unit admission, and mortality [23,25]. The reality of greater risk among those with lower SES may explain greater depression and anxiety in this group; however, it is unclear whether these differences would be perceptible in the earliest days of the pandemic in a sample of adolescents of lower SES.

Research shows that, during COVID-19, being a member of a historically marginalized racial or ethnic group was associated with suicidality [26] and depression [12]. The observed decline in mental health may have been due to the disproportionately negative impact of COVID-19 on racial and ethnic minority communities. Members of racial and ethnic minority groups are more likely to face structural inequities, including inadequate access to health care and overcrowded schools and neighborhoods, both of which contribute to COVID-19 transmission and mortality [25,27]. Although racial and ethnic minority status is closely related to SES in the United States [25], Kim and colleagues [28] attempted to disaggregate the effects of minority status from the effects of SES on COVID-19 outcomes. These authors found that racial minority status, independent of SES, predicted COVID-19 death rate [28]. This may be attributable to the intra- and intergenerational effects of systemic racism on African American individuals’ health and quality of health care.

Daily activities such as part-time work, school, extracurricular activities, and exercise are a source of social support for many adolescents. The sudden removal of these activities —and therefore, of social support—may have resulted in depression and anxiety among US adolescents. Research on Canadian adolescents has shown that having schoolwork to do during quarantine was associated with less depression, while exercise was associated with less loneliness [29]. Research on adolescent athletes shows that athletes involved in team sports experienced more depression during COVID-19 than their peers involved in individual sports, suggesting that the sudden absence of the social component may have contributed to poor mental health [21]. In 2020, Loades and colleagues [30] conducted a rapid systematic review of pre–COVID-19 studies in an attempt to predict the likely effects of social isolation and loneliness on the mental health of adolescents. Based on their sample of articles, these researchers predicted that social isolation due to COVID-19 would result in increased depression among previously healthy adolescents.

This Study
In summary, more research is needed to determine whether American adolescents experienced greater anxiety and depression in the early days of the COVID-19 pandemic; whether this effect was more pronounced among female,
socioeconomically disadvantaged, or racial and ethnic minority adolescents; and whether changes to daily activity were associated with negative mental health. This study contrasts the mental health of US adolescents surveyed prior to the pandemic with that of adolescents surveyed 2 weeks after many states had enacted statewide school closures [2]. Our hypotheses were as follows:

1. Participants surveyed during the pandemic will report greater depression and anxiety than participants surveyed prior to the pandemic.
2. Adolescents surveyed during the pandemic who identify as (A) female, (B) having lower SES, or (C) a racial or ethnic minority will report greater depression and anxiety than their peers surveyed during the pandemic.
3. Adolescents who were prevented from attending (A) work, (B) school, (C) extracurricular activities, or (D) exercise due to the pandemic will report greater depression and anxiety than their peers who were not prevented from carrying out these activities.

It is vital to test these hypotheses as a key step to understanding and designing interventions for adolescents experiencing negative mental health outcomes during and in the aftermath of the COVID-19 pandemic.

Methods

Overview

This 2-group cross-sectional study was a subproject of a larger study. All surveys were taken online. The first group of participants completed the surveys between October 2019 and February 2020, while the second group of responses were received from March 31 to April 3, 2020. This study was approved by the (blinded for review) Institutional Review Board.

Recruitment

Participants

Eligible participants were aged 13-17 years, lived in the United States, and spoke English. The aims of the larger project included evaluating different recruitment approaches for participation in ecological momentary assessment research. Thus, potential subjects for this subproject were recruited using 2 recruitment methods, including Qualtrics and Facebook advertisements. The prepandemic group included participants recruited via Qualtrics and Facebook, while the pandemic group included participants recruited via Facebook only.

Facebook Recruitment

Participants were recruited via paid advertisements posted on Facebook. Advertisements were targeted at the parents of teenagers aged 13-17 years. Parents who clicked on the advertisement were invited to read the informed consent document. Those who provided consent were invited to pass the device to their child, who would then provide assent and complete the eligibility screening. Parents whose children were not with them to provide assent were prompted to enter their child’s phone number; the child would then receive a link to read the assent document and complete the eligibility screening.

Qualtrics Recruitment

Qualtrics is a service that partners with survey research companies to recruit research participants meeting specified inclusion criteria [31]. Qualtrics samples have been shown to be politically and demographically representative of the US population and can replicate known population-level effects [32,33]. Qualtrics sends survey invitations to eligible participants, and compensates participants with rewards, including gift cards, retail points, and airline miles. In this study, Qualtrics sent invitations to the parents of teenagers aged 13-17 years. Only parents who indicated that their children were currently with them and available to provide assent were allowed to complete the consent document. After providing consent, parents were prompted to pass the device to their child so that their child could provide assent and complete the eligibility screening.

Survey Measures

Demographic Information

Demographic questions were asked at the time of enrollment, and included participants’ age, gender, race and ethnicity, and SES. For age, respondents could select a number between 13 and 17. For gender, response options were “female,” “male,” “nonbinary gender,” “female to male transgender,” “male to female transgender,” and “other,” with the option to write in a response. For race and ethnicity, participants could select 1 or more of the following options: “White,” “Black/African American,” “Asian or Pacific Islander,” “American Indian or Alaska Native,” “Hispanic/Latino,” or “other,” with the option to write in a response. To assess SES, participants were asked whether they or their sibling were on free or reduced lunch, and response options were “yes,” “no,” and “I don’t know.”

Mental Health Outcomes

Depression

The 8-item Patient Health Questionnaire (PHQ-8) was used to assess symptoms of depression [34]. Participants were asked to rate the extent to which they had experienced certain symptoms, from “not at all” to “nearly every day.” Example symptoms include “poor appetite or overeating,” “feeling tired or having little energy,” and “feeling down, depressed, or hopeless.” Cutoff scores of 5, 10, and 15 were used to indicate mild, moderate, and severe depression, respectively. Internal consistency for the PHQ-8 is excellent (Cronbach α=.89) [35,36].

Anxiety

The 7-item Generalized Anxiety Disorder (GAD-7) scale was used to assess symptoms of anxiety [37]. Similar to the PHQ-8, participants were asked to rate the extent to which they had experienced certain symptoms, from “not at all” to “nearly every day.” Example symptoms include “not being able to stop or control worrying” and “feeling afraid as if something awful might happen.” As with the PHQ-8, cutoff scores of 5, 10, and 15 were used to indicate mild, moderate, and severe anxiety, respectively [37]. Internal consistency for the GAD-7 is excellent (Cronbach α=.92) [37].
Changes in Daily Activity Due to the Pandemic

A total of 4 pandemic-specific daily activity questions were asked only of the pandemic group. Participants were asked to indicate whether, at present, they were prevented from (1) work, (2) school, (3) extracurricular activities, or (4) exercise due to the pandemic. These were categorical variables, and for each question participants could select 1 of 4 responses: “yes,” “no,” “does not apply,” or “other,” with the option to write in a response.

Procedure

All participants received a link by SMS text messaging to complete the survey. Participants had 3 days after receiving the text message to complete the survey, and participant responses were retained if they completed more than 50% of it. Participants recruited through Qualtrics were compensated through Qualtrics, and those recruited through Facebook were compensated for participation with a check sent to their mailing address.

Analysis

Demographic Variables

Descriptive statistics were used to summarize participants’ age, race and ethnicity, and gender, as well as the percentage of participants who were out of work, school, extracurricular activities, or exercise due to the pandemic. Age was treated as a continuous variable. Given the small number of participants endorsing certain races and ethnicities (eg, Hispanic/Latino, Asian, Pacific Islander), respondents were categorized as non-Hispanic White, non-Hispanic Black, multirace, or other. Due to the small number of participants endorsing genders other than male or female, respondents were categorized as “male,” “female,” or “other.” Participants who indicated that they did not know if they or their siblings were on free or reduced lunch were categorized as not being on free or reduced lunch. Most participants reported being out of school due to the pandemic; thus, the answers “no,” “does not apply,” and “other” were combined in the multivariate analyses. To evaluate demographic differences between the prepandemic and pandemic groups, the Fisher exact test was used for categorical variables, a 2-sample t test was used for normally distributed continuous variables, and a Wilcoxon rank-sum nonparametric test was used for nonnormally distributed continuous variables.

Main Analyses

Statistical analyses were conducted using SAS software, version 9.4 (SAS Institute Inc) [38]. All reported P values were 2-sided, and P<.05 was used to define statistical significance. Hypothesis 1 predicted that being in the pandemic group would be associated with significantly higher depression and anxiety scores compared to the prepandemic group. Analysis of covariance (ANCOVA) was used to test this hypothesis. Anxiety and depression were entered as continuous dependent variables, rather than categorical ones, in order to maximize sensitivity. Group was entered as an independent variable, and demographic variables (age, race, gender, SES) were entered as covariates.

Hypothesis 2 stated that, within the pandemic group, participants identifying as female, as having a lower SES, or as a racial or ethnic minority would be associated with higher depression and anxiety scores. Hypothesis 3 stated that, within the pandemic group, being out of work, school, extracurricular activities, or exercise due to the pandemic would be associated with higher depression and anxiety scores. ANCOVA was again used to test these hypotheses. Anxiety and depression were entered as continuous dependent variables. Work, school, extracurricular activities, exercise, and demographic variables were entered as covariates.

We conducted a sensitivity analysis by excluding all Qualtrics respondents and rerunning the analyses. Furthermore, we evaluated interaction effects between the groups (pre–COVID-19 or COVID-19) and each demographic variable (age, gender, race and ethnicity, and SES) to examine whether further subgroup analysis by demographic characteristics was required.

Results

Participants

There were 100 participants in the prepandemic group, and 134 participants in the pandemic group, for a combined total of 234 participants. Demographic differences between the 2 groups are recorded in Table 1. The samples showed a similar distribution of gender, although statistically significant differences were noted in race, age, and the number of participants on free or reduced lunch. Within the prepandemic group, the mean PHQ-8 and GAD-7 scores were 5.27 (SD 5.06) and 4.67 (SD 5.22), respectively. In the pandemic group, the mean PHQ-8 and GAD-7 scores were 6.81 (SD 5.67) and 5.37 (SD 5.18), respectively. Regarding changes in daily activity during the pandemic, 94% (126/134) of participants reported that they were prevented from attending school during the pandemic, and 85.8% (115/134) were prevented from extracurricular activities. The percentage of participants out of work, school, extracurricular activities, or exercise due to the pandemic is shown in Table 2.
Table 1. Demographic information for prepandemic and pandemic samples (N=234).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Prepandemic (n=100)</th>
<th>Pandemic (n=134)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>14.7 (1.3)</td>
<td>15.2 (1.4)</td>
<td>.01a</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td>.49b</td>
</tr>
<tr>
<td>Female</td>
<td>66 (66)</td>
<td>92 (68.7)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>34 (34)</td>
<td>40 (29.9)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0 (0)</td>
<td>2 (0.01)</td>
<td></td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001b</td>
</tr>
<tr>
<td>White, non-Hispanic</td>
<td>56 (56)</td>
<td>103 (76.9)</td>
<td></td>
</tr>
<tr>
<td>Black, non-Hispanic</td>
<td>18 (18)</td>
<td>8 (6)</td>
<td></td>
</tr>
<tr>
<td>Multirace</td>
<td>11 (11)</td>
<td>12 (9)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>15 (15)</td>
<td>11 (8.2)</td>
<td></td>
</tr>
<tr>
<td>Qualified for free or reduced lunch, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001b</td>
</tr>
<tr>
<td>Yes</td>
<td>51 (51)</td>
<td>41 (30.6)</td>
<td></td>
</tr>
<tr>
<td>No/don’t know</td>
<td>49 (49)</td>
<td>93 (69.4)</td>
<td></td>
</tr>
</tbody>
</table>

aP value from 2-sample t test or Wilcoxon nonparametric test.
bP value from the Fisher exact test.

Table 2. Participants out of work, school, extracurricular activities, or exercise due to the COVID-19 pandemic 2 weeks after nationwide school closures (n=134).

<table>
<thead>
<tr>
<th>Activity</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>31 (23.1)</td>
</tr>
<tr>
<td>No</td>
<td>5 (3.7)</td>
</tr>
<tr>
<td>Does not apply</td>
<td>98 (73.1)</td>
</tr>
<tr>
<td>Other</td>
<td>0 (0)</td>
</tr>
<tr>
<td>School, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>126 (94.0)</td>
</tr>
<tr>
<td>No</td>
<td>4 (3.0)</td>
</tr>
<tr>
<td>Does not apply</td>
<td>2 (1.5)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (1.5)</td>
</tr>
<tr>
<td>Extracurricular activities, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>115 (85.8)</td>
</tr>
<tr>
<td>No</td>
<td>4 (3.0)</td>
</tr>
<tr>
<td>Does not apply</td>
<td>15 (11.2)</td>
</tr>
<tr>
<td>Other</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Exercise, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>76 (57.1)</td>
</tr>
<tr>
<td>No</td>
<td>28 (21.1)</td>
</tr>
<tr>
<td>Does not apply</td>
<td>20 (15.0)</td>
</tr>
<tr>
<td>Other</td>
<td>9 (6.8)</td>
</tr>
</tbody>
</table>
Hypothesis Testing

Hypothesis 1 stated that being in the pandemic group would be associated with significantly higher depression and anxiety scores. In the adjusted model for depression, the pandemic group exhibited higher depression, with a least-squares adjusted mean of 7.62 (SD 1.36) compared to 6.28 (SD 1.42) in the prepandemic group; however, this difference was statistically insignificant ($P=0.08$). Across both groups, female participants had greater depression scores than males (least-squares adjusted means 6.45, SD 0.55 vs 4.70, SD 0.69; $P=0.02$). For anxiety, there was no significant difference in anxiety scores between the pandemic and prepandemic groups (least-squares adjusted means 5.52, SD 1.30 vs 5.01, SD 1.36; $P=0.48$). Again, across both groups, females reported greater anxiety than males (least-squares adjusted means 5.59, SD 0.53 vs 3.58, SD 0.66; $P=0.01$). No other demographic variables were significant predictors of depression or anxiety.

Hypothesis 2 stated that, within the pandemic group, identifying as female, as having a lower SES, or as a racial or ethnic minority would be associated with greater depression and anxiety. In the adjusted model for depression, none of these variables were significant predictors of depression. In the adjusted model for anxiety, being eligible for free or reduced lunch was associated with greater anxiety, such that those in the pandemic group who qualified for free or reduced lunch were more anxious than their peers who did not (least-squares adjusted means 11.17, SD 2.34 vs 8.66, SD 2.16; $P=0.02$).

Hypothesis 3 stated that, within the pandemic group, being out of work, school, extracurricular activities, or exercise due to the pandemic would be associated with higher depression and anxiety scores. No significant associations were detected.

When we conducted a sensitivity analysis by excluding the Qualtrics participants and rerunning the analyses, the results did not change. We also did not detect any significant interaction effect between the groups (pre–COVID-19 or COVID-19) and the demographic characteristics. Therefore, no additional subgroup analyses by demographic characteristics were conducted.

Discussion

Principal Findings

This study examined differences in mental health between adolescents surveyed prior to the COVID-19 pandemic and those surveyed approximately 2 weeks after many states had enacted statewide school closures. Contrary to our hypothesis, adolescents in the pandemic group did not score significantly higher in depression and anxiety after adjusting for covariates. Within the pandemic group, lower SES was associated with higher anxiety. The majority of adolescents were prevented from attending school or participating in extracurricular activities and exercise due to the pandemic, and these changes in daily activity were not associated with depression or anxiety. These findings have important implications for future research on the adolescent experience of the COVID-19 pandemic.

First, we expected to find higher depression and anxiety scores in the pandemic group. This potential finding would have been consistent with existing international literature on COVID-19, which has shown that adolescents and young adults experienced elevated depression and anxiety due to the pandemic [8-10]. In our study, after controlling for the effects of demographic variables, being in the pandemic group was not associated with higher depression or anxiety. It may be that, after only 2 weeks in quarantine, American adolescents did not experience the pandemic as depressing or anxiety-inducing. They may have thought that the restrictions would quickly be lifted or, like many, may not have foreseen the gravity of the virus. Research completed several months into quarantine could yield different results. However, in both the prepandemic and pandemic groups, being female was associated with greater PHQ-8 and GAD-7 scores. This is consistent with existing literature, showing that adolescent females are more at risk for depression and anxiety [22].

In the group of participants surveyed during the pandemic, neither gender nor race was associated with depression or anxiety, which is inconsistent with previously cited research on gender [8,20,21] and race [25-28] during COVID-19. However, SES was associated with anxiety, such that participants of lower SES reported greater anxiety. This is consistent with existing research, which showed that SES was a risk factor for worse mental health outcomes during the pandemic [12,21,39]. It may be that COVID-19 magnified disparities that were not statistically significant in the prepandemic sample. Families of lower SES encountered more COVID-19 infections and fatalities, which may directly challenge the mental health of both infected people and their loved ones [23,25,40]. Underresourced adolescents, in particular, often rely on school for affordable food and school-based health care [41], and may have experienced greater anxiety when those resources were no longer available. In the wake of COVID-19, interventions should focus on adolescents of lower SES as their well-being is uniquely at risk due to the pandemic.

Our results showed that 2 weeks after many school closures, 94% of participants were prevented from going to school due to the pandemic. The majority of participants were prevented from partaking in extracurricular activities or exercise due to the pandemic, perhaps because these opportunities are tied to school for many adolescents. However, these disruptions to daily life within the pandemic group were not associated with higher depression and anxiety scores. This may have occurred for multiple reasons. First, only 2 weeks into quarantine, the absence of daily activities may have felt like a welcome reprieve, especially for adolescents with social anxiety [42], as well as students who experience bullying [43] or discrimination [44,45] in the context of their daily activities. Second, adolescents in this study may have experienced more opportunities for prosocial behavior as a result of the pandemic, such as helping, caring for, and comforting friends and family members; prosocial behavior is associated with positive adjustment [46]. Last, adolescents have a demonstrated ability to use technology to navigate developmental and mental health challenges; in our study, adolescents may have found entertainment, support, and distraction through digital interaction, such that any effects of COVID-19 on mental health were mitigated [47].

https://formative.jmir.org/2021/12/e30702
Limitations
This study was not without limitations. First, the pre-pandemic and pandemic groups were significantly different in terms of age, race and ethnicity, and qualifying for free or reduced lunch. Additionally, pre-pandemic data were collected in the fall and winter months, while pandemic data was collected in the spring. This may have obscured differences caused by the pandemic. However, with the exception of season, ANCOVA should have controlled for all group differences. Second, the majority of participants did not work, making it impossible to ascertain the effects of the pandemic on adolescents who lost employment. Third, there may be constructs better suited than “anxiety” and “depression” for understanding adolescents’ experience of the pandemic. Negative experiences such as fear, boredom, or confusion may be more appropriate, as might positive experiences such as feeling less social or academic pressure, or feeling more supportive and supported online and offline. Further, the 4 pandemic-specific daily activity questions were not based on any validated scales, so these results should be interpreted with caution. Last, because data were collected in early April, findings may not be generalizable to the later periods of the pandemic.

Conclusion
This research suggested that, at least in the early days of the pandemic, adolescents had not developed significantly greater anxiety or depression at the population level. Further, 2 weeks after many school closures, reductions in daily activities were not associated with anxiety and depression. However, we found that adolescents from lower socioeconomic backgrounds experienced significantly more anxiety during the pandemic than their peers of higher SES. Both instrumental and mental health interventions for adolescents of lower SES are imperative. More research is needed to understand the trajectory of adolescents’ mental health experiences during the COVID-19 pandemic, as well as the long-term impact of the pandemic on adolescent mental health.

Acknowledgments
This work was supported by a collaborative research agreement with Facebook. We would like to thank the adolescents who participated in this study.

Conflicts of Interest
None declared.

References


Abbreviations

ANOVA: analysis of covariance
GAD-7: 7-item Generalized Anxiety Disorder
PHQ-8: 8-item Patient Health Questionnaire
SES: socioeconomic status

©Anna Jolliff, Qianqian Zhao, Jens Eickhoff, Megan Moreno. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/).
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.
Pandemic Information Dissemination and Its Associations With the Symptoms of Mental Distress During the COVID-19 Pandemic: Cross-sectional Study

Ole Myklebust Amundsen¹; Asle Hoffart²,³, ClinPsy, PsyD; Sverre Urnes Johnson²,³, ClinPsy, PsyD; Omid V Ebrahimí²,³, ClinPsy

¹Department of Clinical Psychology, University of Bergen, Bergen, Norway
²Modum Bad Psychiatric Hospital, Vikersund, Norway
³Department of Psychology, University of Oslo, Oslo, Norway

Corresponding Author:
Ole Myklebust Amundsen
Department of Clinical Psychology
University of Bergen
Årstadveien 17
Bergen, 5009
Norway
Phone: 47 45209507
Email: oleamundsen@me.com

Abstract

Background: The 2020-2021 COVID-19 pandemic has added to the mental health strain on individuals and groups across the world in a variety of ways. Viral mitigation protocols and viral spread affect people on all continents every day, but at widely different degrees. To understand more about the mental health consequences of the pandemic, it is important to investigate whether or how people gather pandemic-related information and how obtaining this information differentially affects individuals.

Objective: This study aimed to investigate whether and to what extent higher levels of COVID-19–related media consumption across information sources are associated with the symptoms of anxiety, health anxiety, and depression, and whether and to what extent using social media and online interactive platforms versus traditional media platforms is associated with the symptoms of anxiety, health anxiety, and depression. Additionally, we aimed to investigate whether and to what extent avoidance of COVID-19–related information is associated with the aforementioned symptoms.

Methods: In a cross-sectional preregistered survey, 4936 participants responded between June 22 and July 13, 2020. Eligible participants were adults currently residing in Norway and were thus subjected to identical viral mitigation protocols. This sample was representative of the Norwegian population after utilizing an iterative raking algorithm to conduct poststratification. As 2 subgroups (transgender and intersex individuals) were too small to be analyzed, the final sample for descriptive statistics and regressions included 4921 participants. Multiple regressions were used to investigate associations between the symptoms of psychopathology and COVID-19–related information dissemination. Part correlations were calculated as measures of the effect size for each predictor variable. Due to the large anticipated sample size, the preregistered criterion for significance was set at $P<.01$.

Results: The symptoms of anxiety and health anxiety were significantly associated with obtaining information from newspapers ($P<.001$), social media ($P<.001$), and the broader categories of online interactive ($P<.001$) and traditional media ($P<.001$). The symptoms of depression were significantly associated with obtaining information from newspapers ($P=.003$), social media ($P=.009$), and the broader category of online interactive media ($P<.001$). Additionally, avoidance of COVID-19–related information showed a significant association in all 3 domains of psychopathological symptoms (anxiety and depression, $P<.001$; health anxiety, $P=.007$).

Conclusions: This study found significant associations between the symptoms of psychopathology and the use of media for obtaining information related to the COVID-19 pandemic. Significant findings for obtaining information through newspapers, social media, and online interactive media were seen across all 3 measures of psychopathology. Avoidance of COVID-19–related information and associations with the symptoms of psychopathology emerged as core findings, with generally higher effect sizes compared with information attainment.
highly engaged in information-seeking behavior tend to become more focused on the threatening aspects of infectious diseases, experiencing increased distress and anxiety [12,13]. This is especially so when individuals perceive themselves as having little control over the threat [13].

Several years of research have given credence to the idea that increased consumption of information from social media in general can be detrimental to psychological health [14–19], and that increased use of such media is associated with the symptoms of both anxiety and depression [17,20–22]. Social media (eg, Twitter, Facebook, TikTok, Snapchat, Instagram, and WeChat), characterized by fast-paced information streams, are often not curated or validated for facts [23] and have in recent years been an arena for the spread of misinformation [24]. In the current pandemic, social media have been used to dispense information more rapidly to the population [25].

Information is also actively spread through other platforms, such as television, newspapers, websites, and online forums. Additionally, individuals are also informed about the pandemic face-to-face via their peers. News distributed on television can be especially volatile. A study found that the median airtime for medical news stories was just 33 seconds, and these stories did not cite the origin of the information they provided and did not convey recommendations to the public [26]. The association of consumption of information obtained via traditional media with mental distress has also been investigated, but the interaction is less clear than it is for social media [27].

Furthermore, the avoidance of news can have direct consequences for individuals and the society at large. In the current infodemic, news avoidance may be a consequence of news overload (ie, unintentional avoidance). During the current pandemic, relevant advice from intergovernmental, national, and regional authorities has been constantly changing, putting a large strain on individuals who want to keep themselves updated on the current recommendations for hygienic behavior and social distancing. An overload of information sources from platforms, such as newspapers and television, may cause those avoiding these platforms to seek pandemic-relevant information from other sources, such as friends and family, which has been found to be related to decreased adherence to social distancing protocols [28]. Thus, understanding and combating avoidance of information, specifically from official sources, is crucial.

Avoiding information can entail not being confronted with uncertainty. In crises, such as the current pandemic, large groups of people are required to live with a far higher degree of everyday uncertainty than usual [29]. Individuals with high intolerance of uncertainty attempt to reduce this uncertainty through behavioral control, exemplified as checking the internet and other sources for information [29–31]. Furthermore, health anxiety is tied to an increase in the search for health-related information, with psychological distress as a consequence [32].
With an increase in the tendency to search for information, there is also an increase in different media that provide the information, with potentially differential effects.

Thus, it is crucial to investigate how dissemination of divergent sources of information is related to mental health symptoms in the general population during the present COVID-19 pandemic and parallel infodemic.

**Research Questions and Hypotheses**

The following research questions and hypotheses were investigated, as presented in the preregistered protocol of the study:

1. Research questions: Is there a differential effect among different information sources on health anxiety, depression, and anxiety? To what extent and how are different information sources related to the symptoms of health anxiety, depression, and anxiety?

2. Hypothesis 1: Media consumption across all information sources will significantly be associated with depression and anxiety symptoms, with increased media consumption in general associated with higher levels of health anxiety, depression, and general anxiety.

3. Hypothesis 2: Using social media and online interactive platforms (ie, forums and blogs) to obtain news about the pandemic in comparison to using traditional media (ie, television, radio, and newspapers) will be associated with higher levels of general anxiety, depression, and health anxiety. Actively staying away from information will further significantly be associated with higher levels of anxiety, depression, and health anxiety.

**Methods**

**Overview**

This cross-sectional study is part of The Norwegian COVID-19 Mental Health and Adherence Project, and utilizes data collected in the second stage of data collection in this project. This study investigated the association of various sources of information acquisition concerning COVID-19 with psychopathology, specifically the symptoms of anxiety, depression, and health anxiety during the COVID-19 pandemic. Following the guidelines of Strengthening the Reporting of Observational Studies in Epidemiology [33], and the health estimate reporting standards laid out in the GATHER statement [34], this study was designed and preregistered prior to any data collection. The preregistered protocol can be found at ClinicalTrials.gov (NCT04442360). All components of the submitted study adhere to its preregistered protocol.

**Study Design, Participants, Procedure, and Timing**

The data for this cross-sectional study were collected via an online survey between June 22 and July 13, 2020, and involved the second stage of data collection for the project. In the first wave of data collection, a survey was distributed on national, regional, and local information platforms (ie, television, radio, and newspapers), in addition to dissemination to a random selection of Norwegian adults through a Facebook Business algorithm. Details regarding procedure and timing can be found elsewhere [9]. In the first stage, data were obtained from 10,061 participants. For the second stage of data collection, all participants who had provided informed consent to participate further were invited to take part. The eligible participants were adults (ie, aged 18 years or above) currently residing in Norway, who were thus subject to identical VMPs. A total of 4936 participants made up the sample for this study. The survey was administered approximately 1 week after major VMPs were lifted in Norway and included a period when the national VMPs and guidelines did not shift. Additionally, no novel information was provided by the Norwegian government regarding social distancing protocols during this period.

**Measurements**

**Demographic Characteristics**

Information regarding participant age, sex, education level, ethnic background, and regional affiliation level were collected.

**Symptoms of Psychopathology**

The Generalized Anxiety Disorder-7 (GAD-7) [35] is a scale for identifying the level of anxiety. It consists of 7 items measuring anxiety on a 4-point Likert scale (range 0-3), with the total score ranging from 0 to 21. For GAD-7, internal consistency was good, with a Cronbach α of .90.

The Patient Health Questionnaire-9 (PHQ-9) [36] is a measure of depression severity that consists of 9 items scored on a 4-point Likert scale (range 0-3), with the total score ranging from 0 to 27. The internal consistency of this scale was excellent in this sample, with a Cronbach α of .91.

The symptoms of health anxiety were measured with 2 items from the validated Health Anxiety Inventory [37] and 2 items adapted for the COVID-19 pandemic, with one item measuring the specific fear of being infected by the coronavirus and another item measuring the fear of dying from the coronavirus on a 4-point Likert scale (range 0-3). Internal consistency was acceptable for these health anxiety–related questions, with a Cronbach α of .77.

**Information-Seeking Behavior**

Participants were asked to estimate the amount of time they had spent obtaining information regarding the COVID-19 pandemic using various media platforms since the beginning of March 2020, which was the onset of pandemic restrictions in Norway. This included (1) recognized newspapers; (2) television channels; (3) social media; (4) forums, blogs, podcasts, and other online outlets (excluding online newspapers); (5) friends, family, and acquaintances; and (6) other sources. Moreover, active avoidance of COVID-19–related information on all media was measured. These variables were measured on an 8-point Likert scale (range 0-7; 0=never, 7=multiple times per hour). Two new variables were created from the existing media variables. Using online interactive media was defined as using social media, forums, and blogs, and was consequently the sum of these variables. Using traditional media was defined as using television and newspapers, and consisted of the sum of these variables.
Statistical Analyses

The descriptive analyses of the present data were reported using means and standard deviations. Two educational groups were collapsed due to a low N value, and these groups were also used in multiple regressions. For hypothesis 1, 3 separate multiple regression analyses with the symptoms of anxiety, depression, and health anxiety as criterion variables were conducted. The predictor variables were the extents of information obtained about the COVID-19 pandemic from (1) recognized newspapers; (2) television channels; (3) social media; (4) forums, blogs, podcasts, and other online outlets (excluding online newspapers); (5) friends, family, and acquaintances; and (6) other sources, as well as actively staying away from information. Additionally, we controlled for age, gender, education level, and the presence of a psychiatric diagnosis.

For hypothesis 2, 3 separate multiple regression analyses were conducted, with the symptoms of anxiety, depression, and health anxiety as criterion variables. Media variables were collapsed as described above into traditional media and online interactive media. Additionally, we controlled for age, gender, education level, and the presence of a psychiatric diagnosis.

For all multiple regressions, part correlations were calculated. Part correlation, also referred to as semipartial correlation, is a measure low in bias that is easily interpretable. Part correlations are estimates of the strength of a predictive relationship and can be interpreted as an effect size measure using the Cohen criteria of small effect size >0.10, medium effect size >0.30, and large effect size >0.50 [38,39]. Standard criteria for multicollinearity were fulfilled [40] in all multiple regression models, and all assumptions were met.

The preregistered criterion for significance was set at $P<.01$, given the anticipated sample size. All statistical analyses were conducted using R (version 4.0.3; The R Project for Statistical Computing).

Poststratification Weights

In this study, the sampled gender, age, education, regional affiliation, and ethnic background deviated somewhat from the population parameters. All these deviations (minor deviations as well as larger deviations) were weighted and adjusted to accurately reflect the Norwegian adult population. More weight was assigned to underrepresented units and less weight to overrepresented units. Weights were calculated using the R packages “anesrake” (version 0.8) and “survey” (version 4.0). These packages use an iterative algorithm (ie, raking ratio estimation) to iteratively assign appropriate weights to each subgroup by turn to avoid the distribution matching of one factor skewing the distribution of other factors. The weights outputted as a result of this algorithm were applied to the data set, resulting in a weighted data set with parameters closely matching that of the true population and yielding a highly representative sample. All statistical analyses utilized this weighted representative sample.

Results

Sample Characteristics

A total of 4936 individuals were included in this sample, but as 2 of the subgroups (transgendered and intersex individuals) were too small to be factors in our analyses, a sample of 4921 individuals was used for descriptive statistics (Table 1) and for all multiple regression analyses. Table 1 also shows the weighted N values after poststratification with the population parameters as a reference point. Means and standard deviations for the media variables and avoidance can be found in Multimedia Appendix 1, showing that mean time spent gathering information about COVID-19 was the highest for newspapers, television, and social media. A correlation matrix for media variables and avoidance can be found in Multimedia Appendix 2, showing a negative correlation of avoidance with all media variables as expected.
Table 1. Sample characteristics and weighted characteristics (N=4921).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sample, n (%)</th>
<th>Weighted, n (%)</th>
<th>Actual population, %</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>3911 (79.48)</td>
<td>2427 (49.32)</td>
<td>49.77</td>
</tr>
<tr>
<td>Male</td>
<td>1010 (20.52)</td>
<td>2494 (50.68)</td>
<td>50.23</td>
</tr>
<tr>
<td><strong>Age group (years)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-30</td>
<td>1703 (34.61)</td>
<td>1069 (21.73)</td>
<td>23.22</td>
</tr>
<tr>
<td>31-44</td>
<td>1606 (32.64)</td>
<td>1340 (27.24)</td>
<td>24.30</td>
</tr>
<tr>
<td>45-64</td>
<td>1344 (27.31)</td>
<td>1707 (34.70)</td>
<td>31.26</td>
</tr>
<tr>
<td>65 or above</td>
<td>268 (5.44)</td>
<td>804 (16.34)</td>
<td>21.22</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Junior high school&lt;sup&gt;a&lt;/sup&gt;</td>
<td>191 (3.88)</td>
<td>573 (11.64)</td>
<td>25.40</td>
</tr>
<tr>
<td>Completed high school</td>
<td>736 (14.96)</td>
<td>1931 (39.25)</td>
<td>37.00</td>
</tr>
<tr>
<td>Currently studying</td>
<td>775 (15.75)</td>
<td>409 (8.32)</td>
<td>6.70</td>
</tr>
<tr>
<td>Completed university degree</td>
<td>3219 (65.41)</td>
<td>2006 (40.78)</td>
<td>30.90</td>
</tr>
<tr>
<td><strong>Ethnic background</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Native Norwegian</td>
<td>4563 (92.73)</td>
<td>4346 (88.32)</td>
<td>85.29</td>
</tr>
<tr>
<td>European</td>
<td>274 (5.57)</td>
<td>401 (8.16)</td>
<td>7.58</td>
</tr>
<tr>
<td>Asian</td>
<td>39 (0.79)</td>
<td>117 (2.38)</td>
<td>4.56</td>
</tr>
<tr>
<td>African</td>
<td>6 (0.12)</td>
<td>18 (0.37)</td>
<td>1.85</td>
</tr>
<tr>
<td>North America/Oceania</td>
<td>15 (0.30)</td>
<td>17 (0.35)</td>
<td>0.27</td>
</tr>
<tr>
<td>South/Middle/Latin-America</td>
<td>24 (0.49)</td>
<td>21 (0.43)</td>
<td>0.45</td>
</tr>
<tr>
<td><strong>Regional affiliation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>East Norway</td>
<td>3103 (63.06)</td>
<td>2966 (60.28)</td>
<td>58.32</td>
</tr>
<tr>
<td>West Norway</td>
<td>1162 (23.61)</td>
<td>922 (18.74)</td>
<td>20.28</td>
</tr>
<tr>
<td>Mid-Norway</td>
<td>482 (9.79)</td>
<td>781 (15.87)</td>
<td>15.95</td>
</tr>
<tr>
<td>Northern Norway</td>
<td>174 (3.54)</td>
<td>251 (5.11)</td>
<td>5.45</td>
</tr>
</tbody>
</table>

<sup>a</sup>This category is collapsed and consists of individuals who did and those who did not complete junior high school.

Information-Seeking Behavior and the Symptoms of Anxiety

The multiple regression model examining the factors associated with the symptoms of anxiety can be found in Table 2 and Multimedia Appendix 3. Multimedia Appendix 3 displays the regression results for the variables online interactive media and traditional media, explaining 33% of the variance in the data. Gender was a significant predictor of anxiety, with female gender being associated with higher levels of anxiety symptoms. Age was also a significant predictor of anxiety symptoms, with lower age being associated with higher symptoms of anxiety. Additionally, having a pre-existing mental health condition was associated with increased anxiety symptoms. Time spent obtaining information about the pandemic using newspapers was a significant predictor of anxiety symptoms, and more time spent reading newspapers was associated with a higher degree of anxiety symptoms.
Table 2. Predictors of anxiety symptoms in the weighted representative sample (N=4921; adjusted $R^2=0.33$).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
<th>SE of beta</th>
<th>$P$ value</th>
<th>Part correlation (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>5.16</td>
<td>0.39</td>
<td>&lt;.001</td>
<td>1.00</td>
</tr>
<tr>
<td>Gender$^a$</td>
<td>−0.59</td>
<td>0.18</td>
<td>.001</td>
<td>−0.06</td>
</tr>
<tr>
<td>Age</td>
<td>−0.06</td>
<td>0.01</td>
<td>&lt;.001</td>
<td>−0.18</td>
</tr>
<tr>
<td>Education</td>
<td>−0.19</td>
<td>0.09</td>
<td>.03</td>
<td>−0.04</td>
</tr>
<tr>
<td>Mental health condition</td>
<td>4.34</td>
<td>0.29</td>
<td>&lt;.001</td>
<td>0.37</td>
</tr>
<tr>
<td>Newspapers</td>
<td>0.28</td>
<td>0.07</td>
<td>&lt;.001</td>
<td>0.08</td>
</tr>
<tr>
<td>Television</td>
<td>0.01</td>
<td>0.07</td>
<td>.86</td>
<td>0.00</td>
</tr>
<tr>
<td>Social media</td>
<td>0.24</td>
<td>0.07</td>
<td>&lt;.001</td>
<td>0.08</td>
</tr>
<tr>
<td>Forums and blogs</td>
<td>0.14</td>
<td>0.09</td>
<td>.10</td>
<td>0.04</td>
</tr>
<tr>
<td>Friends and family</td>
<td>−0.06</td>
<td>0.10</td>
<td>.58</td>
<td>−0.01</td>
</tr>
<tr>
<td>Other</td>
<td>0.04</td>
<td>0.08</td>
<td>.57</td>
<td>0.01</td>
</tr>
<tr>
<td>Avoidance</td>
<td>0.36</td>
<td>0.08</td>
<td>&lt;.001</td>
<td>0.10</td>
</tr>
</tbody>
</table>

$^a$Female=0; male=1.

Time spent obtaining information from social media was a significant predictor of anxiety symptoms, with more time spent being associated with a higher degree of anxiety symptoms. Obtaining information using both traditional media and online interactive media was a significant predictor of anxiety, and the relationship was positive. Finally, avoiding information about COVID-19 entirely was a significant predictor of anxiety symptoms, with a higher degree of avoidance being associated with an increased level of anxiety symptoms.

Among these significant associations, part correlations revealed the factors most strongly associated with the symptoms of anxiety while accounting for all other variables. The most prominent factors associated with the symptoms of anxiety included having a pre-existing mental health condition (part correlation=0.37), age (part correlation=−0.18), obtaining information from online interactive media (part correlation=0.10), and avoiding information about the pandemic (part correlation=0.10), all revealing small to medium effect sizes. Lower effect sizes emerged for obtaining information concerning the COVID-19 pandemic from social media (part correlation=0.08), for obtaining information from newspapers (part correlation=0.08), and generally for traditional media (part correlation=0.08) and the effect of gender (part correlation=−0.06).

Obtaining information about the pandemic from television, forums and blogs, family and friends, and other sources was unrelated to the symptoms of anxiety. Education level was also unrelated to the symptoms of anxiety at our significance level.

**Information-Seeking Behavior and the Symptoms of Depression**

The multiple regression model examining the factors associated with depressive symptoms is depicted in Table 3 and Multimedia Appendix 4. Multimedia Appendix 4 displays the regression results for the variables online interactive media and traditional media, explaining 36% of the variance in the data. Age, gender, and education level were significant predictors of depressive symptoms, where lower age, lower education level, and a trend toward female gender were associated with a higher level of depressive symptoms. Having a pre-existing mental health condition was associated with a greater level of depressive symptoms. Time spent obtaining information about the pandemic from newspapers and social media was a significant predictor of depressive symptoms, with more time spent reading newspapers being associated with a higher level of depressive symptoms.
### Table 3. Predictors of depressive symptoms in the weighted representative sample (N=4921; adjusted $R^2=0.33$).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
<th>SE of beta</th>
<th>$P$ value</th>
<th>Part correlation (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>8.42</td>
<td>0.54</td>
<td>&lt;.001</td>
<td>1.00</td>
</tr>
<tr>
<td>Gender(^a)</td>
<td>−0.68</td>
<td>0.25</td>
<td>.005</td>
<td>−0.05</td>
</tr>
<tr>
<td>Age</td>
<td>−0.07</td>
<td>0.01</td>
<td>&lt;.001</td>
<td>−0.16</td>
</tr>
<tr>
<td>Education</td>
<td>−0.34</td>
<td>0.12</td>
<td>.004</td>
<td>−0.05</td>
</tr>
<tr>
<td>Mental health condition</td>
<td>6.41</td>
<td>0.39</td>
<td>&lt;.001</td>
<td>0.41</td>
</tr>
<tr>
<td>Newspapers</td>
<td>0.27</td>
<td>0.09</td>
<td>.002</td>
<td>0.06</td>
</tr>
<tr>
<td>Television</td>
<td>−0.12</td>
<td>0.09</td>
<td>.19</td>
<td>−0.03</td>
</tr>
<tr>
<td>Social media</td>
<td>0.22</td>
<td>0.08</td>
<td>.009</td>
<td>0.05</td>
</tr>
<tr>
<td>Forums and blogs</td>
<td>0.18</td>
<td>0.11</td>
<td>.11</td>
<td>0.03</td>
</tr>
<tr>
<td>Friends and family</td>
<td>−0.25</td>
<td>0.13</td>
<td>.06</td>
<td>−0.04</td>
</tr>
<tr>
<td>Other</td>
<td>−0.02</td>
<td>0.10</td>
<td>.82</td>
<td>0.001</td>
</tr>
<tr>
<td>Avoidance</td>
<td>0.64</td>
<td>0.11</td>
<td>&lt;.001</td>
<td>0.13</td>
</tr>
</tbody>
</table>

\(^a\)Female=0; male=1.

Obtaining information about the pandemic through online interactive media was a significant predictor of depressive symptoms, where more time spent was associated with a higher level of depressive symptoms. Avoiding information entirely was a significant predictor of depressive symptoms, where increasing avoidance of COVID-19–related information was associated with a higher level of depressive symptoms. Among these significant associations, part correlations revealed the factors most strongly associated with depressive symptoms while accounting for all other variables. The strongest factors associated with depressive symptoms included having a pre-existing mental health condition (part correlation=0.41), age (part correlation=−0.16), and avoiding information about COVID-19 (part correlation=0.13), all revealing small to medium effect sizes. Smaller effect sizes were observed for education level (part correlation=−0.06), and obtaining news about the pandemic through newspapers (part correlation=0.06), online interactive media (part correlation=0.08), and traditional media (part correlation=0.03).

Obtaining information about COVID-19 from television, forums and blogs, family and friends, and other sources was unrelated to depressive symptoms. Obtaining information from traditional media was also unrelated to depressive symptoms.

**Information-Seeking Behavior and the Symptoms of Health Anxiety**

The multiple regression model examining the factors associated with the symptoms of health anxiety can be found in Table 4 and Multimedia Appendix 5. Multimedia Appendix 5 displays the regression results for the variables online interactive media and traditional media, and the regressions in Table 4 and Multimedia Appendix 5 explain 16% and 15% of the variance in the data, respectively. Education was a significant predictor of health anxiety symptoms, with a lower education level being associated with a higher level of health anxiety symptoms. Having a pre-existing mental health diagnosis was a significant predictor of health anxiety symptoms.
Table 4. Predictors of health anxiety symptoms in the weighted representative sample (N=4921; adjusted $R^2=0.16$).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
<th>SE of beta</th>
<th>P value</th>
<th>Part correlation (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.52</td>
<td>0.18</td>
<td>.003</td>
<td>1.00</td>
</tr>
<tr>
<td>Gendera</td>
<td>−0.13</td>
<td>0.09</td>
<td>.13</td>
<td>−0.03</td>
</tr>
<tr>
<td>Age</td>
<td>0.00</td>
<td>0.00</td>
<td>&lt;.001</td>
<td>0.03</td>
</tr>
<tr>
<td>Education</td>
<td>−0.19</td>
<td>0.04</td>
<td>&lt;.001</td>
<td>−0.09</td>
</tr>
<tr>
<td>Mental health condition</td>
<td>1.19</td>
<td>0.14</td>
<td>&lt;.001</td>
<td>0.24</td>
</tr>
<tr>
<td>Newspapers</td>
<td>0.20</td>
<td>0.03</td>
<td>&lt;.001</td>
<td>0.13</td>
</tr>
<tr>
<td>Television</td>
<td>0.02</td>
<td>0.03</td>
<td>.61</td>
<td>0.01</td>
</tr>
<tr>
<td>Social media</td>
<td>0.12</td>
<td>0.03</td>
<td>&lt;.001</td>
<td>0.09</td>
</tr>
<tr>
<td>Forums and blogs</td>
<td>0.03</td>
<td>0.05</td>
<td>.49</td>
<td>0.02</td>
</tr>
<tr>
<td>Friends and family</td>
<td>−0.03</td>
<td>0.05</td>
<td>.61</td>
<td>−0.01</td>
</tr>
<tr>
<td>Other</td>
<td>0.02</td>
<td>0.04</td>
<td>.69</td>
<td>0.01</td>
</tr>
<tr>
<td>Avoidance</td>
<td>0.09</td>
<td>0.04</td>
<td>.007</td>
<td>0.06</td>
</tr>
</tbody>
</table>

*aFemale=0; male=1.

Time spent obtaining information from newspapers was a significant predictor of health anxiety symptoms, with more time spent being associated with an increased level of health anxiety symptoms. Time spent obtaining information from social media was a significant predictor of health anxiety symptoms, with more time spent being associated with a higher level of health anxiety symptoms. Obtaining information about the pandemic from both traditional media and online interactive media was a significant predictor of health anxiety, where more time spent on these media was associated with increased symptoms. Avoiding information entirely was a significant predictor of health anxiety symptoms, with higher levels of avoidance being positively associated with health anxiety symptoms.

Among these significant associations, part correlations revealed the factors most strongly associated with the symptoms of health anxiety while accounting for all other variables. The strongest factors associated with the symptoms of health anxiety included having a pre-existing mental health condition (part correlation=0.23), using traditional media to obtain information about the pandemic (part correlation=0.14), and time spent gathering information from newspapers (part correlation=0.13), in addition to using online interactive media in general (part correlation=0.10), revealing small effect sizes. Obtaining information through social media (part correlation=0.09), education level (part correlation=0.09), and avoiding news about COVID-19 (part correlation=0.06) had smaller effect sizes.

Age; gender; and information attainment related to COVID-19 from television, family and friends, forums and blogs, and other sources were unrelated to the symptoms of health anxiety.

Discussion

Hypothesized Effects

This study investigated the association of various sources of pandemic information dissemination with psychopathology during the COVID-19 pandemic in a Norwegian sample. For hypothesis 1, the strongest effect sizes were found for using social media and newspapers to obtain information about COVID-19. Support was not found for the hypothesized associations of obtaining information from television, forums and blogs, friends and family, and other sources with the symptoms of anxiety, health anxiety, and depression.

For hypothesis 2, using either online interactive media or traditional media had the strongest effect sizes pertaining to the symptoms of psychopathology, except for the measure of depressive symptoms, where using traditional media was not statistically significant. Additionally, as hypothesized, avoiding news about COVID-19 was significantly associated with the symptoms of anxiety, health anxiety, and depression.

Information-Seeking Behavior and the Symptoms of Psychopathology

Our findings seem to be in line with previous pandemic research results, where the use of different sources of information about COVID-19 was associated with increased symptoms of anxiety, depression, and health anxiety [12,13,41,42]. During such pandemics, individuals are required to live life with a certain degree of uncertainty due to the novelty of the situation [29], such as living with the risk of getting infected, not knowing if others you interact with are contagious, and being constantly wary of potential situations that could put one at risk.

Individuals with high intolerance to uncertainty may attempt to reduce this uncertainty through behaviors, such as checking the internet (ie, social media, forums, and blogs) and other sources (ie, newspapers and television) for information [29-31], exuding internal control over an external problem. Furthermore, health anxiety is tied to an increase in the search for health-related information, with psychological distress as a potential consequence [32]. With an increased tendency to search for information, there is also an increase in different media providing that information, with potentially differential effects.
Additionally, using social media in general is associated with the symptoms of mental distress [14-19].

As with social media, other unverified platforms have been investigated in previous pandemics. During the 2009 swine flu outbreak, language in blogs mentioning the swine flu was more dramatic compared to a control, and this language later spread to newspapers [43], indicating that monitoring of multiple types of media may result in exposure to similar dramatic language over time across different media. Interestingly, a study by Blakey [44] during the 2015 Zika outbreak found that knowledge of the virus predicted anxiety, indicating that it may not only be dramatic language that has a detrimental effect, but also general knowledge about the outbreak itself, an important point that national and regional governments need to take into consideration when preparing information for their citizens.

**Information Avoidance and the Symptoms of Psychopathology**

In line with the findings of this investigation, avoidance in anxiety literature is empirically documented as a maladaptive behavior that maintains anxiety levels [45-50]. Our results indicate that the association of news avoidance with anxiety and depression was greater than the association of the use of different media platforms for information about the pandemic with the symptoms of psychopathology, when viewing part correlations as effect sizes.

One reason for this might be that individuals who explicitly avoid news show general tendencies of avoidance, although existing mental health conditions were controlled for in our analyses. Previous research has shown that those who avoid news intentionally are exposed to news regardless [51-53]. Actively avoiding news and then being exposed to the news regardless might trigger the very feelings one seeks to avoid, although in an uncontrolled context, it might be more hurtful than being exposed to the news in the first place. Avoiding news might be maladaptive in the same sense that avoiding a feared stimuli for a person who has an anxiety disorder hinders recovery. Avoiding such information can shield the individual from information that might trigger existing fearful thoughts, but can possibly increase the span of attentiveness toward general negative experience, which may in turn be reinterpreted as threatening symptom stimuli through attentional biases [54].

Consequently, a core question is how we can reach those avoiding pandemic-related information, as these individuals seem to be at increased risk of experiencing psychopathological symptoms. To better understand the reasons underlying avoidance, future studies should explore whether pandemic-related news avoidance is intentional or unintentional, and how it is related to “news overload.” Skovgaard and Andersen [55] argue that one potential cause for avoidance is the overload of information, which is in accordance with the term infodemic. This connection is further elucidated in recent studies on social media [56,57], showing that news avoidance directly affects news efficacy.

Furthermore, individuals may be engaged in intentional avoidance due to low trust in news and the negativity of news [55]. The authors propose that these problems underlying intentional avoidance may be solved by providing clear and concise information, avoiding presentation of opinions, sticking to reporting facts, and rebalancing negative perspectives with future-oriented solutions to the presented problems. Rebalancing the overload of negative information present during pandemics with solutions on how to tackle the pandemic is also important from a psychological perspective, with the ability to increase the psychological need of competence and self-efficacy [58].

Governments might also use the research on incidental exposure to news to position information on platforms and in contexts where the information is likely to be seen, even in passing, especially considering that there are groups of people who use social media as their primary source of news [41]. As both this study and other studies have shown associations of social media use for news consumption regarding the current pandemic and social media use in general with the symptoms of psychopathology, it becomes vitally important to engage readers to both seek and receive fact-checked information on these media. Another possible solution could be attempted by presenting imperative pandemic-related news along with entertainment programs on television, and in the context of entertainment on the internet, previous studies have shown that people are exposed to this information even in passing [41]. Through these strategies designed to tackle intentional and unintentional avoidance, journalists and health-policy makers may provide avoidant individuals with essential life-saving information, and ease added symptoms of psychopathology associated with avoidance of news related to the pandemic.

**Strengths and Limitations**

A major strength of this study is that data collection was conducted in a period where pandemic restrictions had just been lifted, and no significant changes to the pandemic protocols in Norway occurred. Although the collected data were not fully representative of the Norwegian population given the oversampling and undersampling of several highlighted subgroups, poststratification weights were calculated and applied to closely match the sample with the true distribution in the population, yielding a representative sample of the Norwegian adult population. The study used a cross-sectional design, which impair the ability to infer causal effects. Another limitation is using self-reporting of symptoms and inquiring about the presence of a psychiatric diagnosis rather than performing diagnostic interviews. As with traditional sampling techniques involving voluntary participation, potential self-selection of participants may have occurred. However, self-selection was accounted for statistically through the described weighting and raking procedures.

**Conclusion**

This study found significant associations between the symptoms of psychopathology and the use of media for obtaining information related to the COVID-19 pandemic. Significant findings for obtaining information through newspapers, social media, and online interactive media were seen across all 3 measures of psychopathology. Avoidance of COVID-19–related information and associations with the symptoms of psychopathology emerged as core findings, with generally higher effect sizes compared with informational attainment. In
the age of social media and fast-paced informational streams, presenting up-to-date curated pandemic information is imperative to mitigate the associated detrimental mental health consequences of information dissemination, and efficacious communication is important to hinder viral spread, keep populations up to date on the latest hygienic recommendations, and counteract the effects of infodemic misinformation and news overload. Effectively and accurately conveying information can directly benefit the mental health of citizens.

Further research should be conducted to elucidate the causal mechanisms underlying the associations between media use and the symptoms of psychopathology, and between news avoidance and the symptoms of psychopathology. Public health officials must direct their efforts toward creating better opportunities for individuals to acquire validated information related to the pandemic, nudging them to trusted sources of information, and preventing news overload.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data for media variables.
[DOCX File, 17 KB - formative_v5i12e28239_app1.docx ]

Multimedia Appendix 2
Correlation matrix for media variables.
[DOCX File, 17 KB - formative_v5i12e28239_app2.docx ]

Multimedia Appendix 3
Predictors of anxiety symptoms in the weighted representative sample.
[DOCX File, 17 KB - formative_v5i12e28239_app3.docx ]

Multimedia Appendix 4
Predictors of depressive symptoms in the weighted representative sample.
[DOCX File, 17 KB - formative_v5i12e28239_app4.docx ]

Multimedia Appendix 5
Predictors of health anxiety symptoms in the weighted representative sample.
[DOCX File, 17 KB - formative_v5i12e28239_app5.docx ]

References


Abbreviations

- **GAD-7**: Generalized Anxiety Disorder-7
- **PHQ-9**: Public Health Questionnaire-9
- **VMP**: viral mitigation protocol
Re-examining COVID-19 Self-Reported Symptom Tracking Programs in the United States: Updated Framework Synthesis

Miranda Lynn Janvrin¹,², MPH; Jessica Korona-Bailey¹,², MPH; Tracey Pérez Koehlmoos², MHA, PhD
¹The Henry M Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD, United States
²Uniformed Services University of the Health Sciences, Bethesda, MD, United States

Corresponding Author:
Miranda Lynn Janvrin, MPH
The Henry M Jackson Foundation for the Advancement of Military Medicine
6720B Rockledge Dr
Suite 100
Bethesda, MD, 20817
United States
Phone: 1 6035403059
Email: miranda.janvrin@gmail.com

Abstract

Background: Early in the pandemic, in 2020, Koehlmoos et al completed a framework synthesis of currently available self-reported symptom tracking programs for COVID-19. This framework described relevant programs, partners and affiliates, funding, responses, platform, and intended audience, among other considerations.

Objective: This study seeks to update the existing framework with the aim of identifying developments in the landscape and highlighting how programs have adapted to changes in pandemic response.

Methods: Our team developed a framework to collate information on current COVID-19 self-reported symptom tracking programs using the “best-fit” framework synthesis approach. All programs from the previous study were included to document changes. New programs were discovered using a Google search for target keywords. The time frame for the search for programs ranged from March 1, 2021, to May 6, 2021.

Results: We screened 33 programs, of which 8 were included in our final framework synthesis. We identified multiple common data elements, including demographic information such as race, age, gender, and affiliation (all were associated with universities, medical schools, or schools of public health). Dissimilarities included questions regarding vaccination status, vaccine hesitancy, adherence to social distancing, COVID-19 testing, and mental health.

Conclusions: At this time, the future of self-reported symptom tracking for COVID-19 is unclear. Some sources have speculated that COVID-19 may become a yearly occurrence much like the flu, and if so, the data that these programs generate is still valuable. However, it is unclear whether the public will maintain the same level of interest in reporting their symptoms on a regular basis if the prevalence of COVID-19 becomes more common.

(JMIR Form Res 2021;5(12):e31271) doi:10.2196/31271

KEYWORDS
COVID-19; coronavirus; framework analysis; information resources; monitoring; patient-reported outcome measures; self-reported; surveillance; symptom tracking; synthesis; digital health

Introduction

Over the past few years, the COVID-19 pandemic has resulted in more than 175 million deaths around the globe, and it has fundamentally changed the lives of millions more, demanding continual innovation and ingenuity. The increasing isolation brought on by the recommended quarantine and social distancing guidelines has led to further reliance on technology for both employment and social interaction [1,2]. Capitalizing on this, researchers in the field of participatory epidemiology developed programs to track individuals’ symptoms and map trends within communities.

These developments were further spurred by the delay in traditional epidemiological surveillance. Although epidemiological surveillance is fundamental in coordinating strategies for detection and prevention [3], in the early stages
of the COVID-19 pandemic, there were many instances where testing was not available or otherwise inaccessible [4]. It was postulated that this delay could be partially assuaged through an uptake in crowdsourced participatory surveillance efforts, based on the successful monitoring of yearly influenza outbreaks [5,6]. Although outbreaks have considerably declined since the advent of the COVID-19 vaccine, the application of COVID-19 symptom tracking technology to the pandemic has afforded researchers ample data both on virus spread and its impact on the individual.

To shed light on these important efforts in the United States, early in the pandemic, in 2020, Koehlmoos et al [7] completed a framework synthesis of currently available self-reported symptom tracking programs for COVID-19—the disease caused by a specific novel coronavirus. They sought to identify programs in the United States that remotely tracked daily symptom fluctuations among participants and mapped disease spread at a community level. This framework described the programs, partners and affiliates, funding, responses, platform, and intended audience, among other considerations. Ideally, this synthesis would raise awareness of programs while identifying gaps and overlaps [7]. This study seeks to update Koehlmoos et al’s framework, with the aim of identifying developments in the landscape and highlighting how programs have adapted to changes in pandemic response.

Methods

Study Design

Our team developed a framework to collate information on current COVID-19 self-reported symptom tracking programs using the “best-fit” framework synthesis approach [8]. The innovative best-fit framework was chosen for its strength, utility, and appropriateness in drawing conclusions for an evolving subject, as it does not require a further process of interpretation by policy makers and practitioners in order for them to inform practice [9]. Information on each program was collected and organized into a table for ease of comparison.

Target Population

The original framework synthesis served to identify programs tracking symptoms associated with COVID-19 in the US population. The updated version includes all programs identified in the prior synthesis, as well as sought to include any programs implemented thereafter. Inclusion and exclusion criteria are described below.

Inclusion Criteria

Programs that aim to capture and geographically collate self-reported symptoms associated with COVID-19 that are available for use in the United States were included for the analysis. For this synthesis, a symptom tracking program is defined as a program that allows individuals to report COVID-19 symptoms, in an effort to identify geographic areas with emerging disease or changes in disease progression.

Exclusion Criteria

Programs that do not track specific symptoms for COVID-19, those that were symptom checkers intended for individual use only, or those that did not target the US population were excluded from the analysis.

Program Identification

All programs from the previous synthesis were included to document changes, if any. New programs were discovered using a Google search for keywords (eg, “symptom trackers covid,” “symptom forwarders coronavirus,” “symptom forwarding covid,” “symptom tracking coronavirus,” “daily symptom tracking covid,” “daily symptom tracking coronavirus,” “self-reporting covid,” “self-reporting coronavirus”). The time frame for the search for programs ranged from March 1, 2021, to May 6, 2021. This 2-month timeline was selected to reflect the state of COVID-19 symptom tracking programs 1 year after our previous synthesis.

Screening Method

Reviewers (JK, MLJ, and TPK) screened programs to determine whether inclusion criteria were met, and if so, they extracted data from program websites using a standardized form. In the event of conflict between reviewers, final determinations were made by TPK. To gather information that may not be available via program websites, program managers were contacted via email.

Synthesis Method

Data related to program characteristics were extracted from all included programs and organized into a table using the tool developed in the previous synthesis to guide data collection and build the framework for analysis. Data were then synthesized to form meaningful statements about the programs.

Results

We identified 33 programs in total, 8 (24%) of which met the inclusion criteria. All 6 programs identified in the last synthesis were included to document changes, if any. Notably, 2 of those 6 programs had been rebranded since our original synthesis, with COVIDNearYou now known as Outbreaks Near Me and COVID Symptom Tracker now known as COVID Symptom Study. Two newly identified programs, COVIDSymptom and COVID Control, met the inclusion criteria. Thus, information was gathered from all 8 eligible programs (ie, BeatCOVID19Now, COVIDcast, COVID Control, COVIDSymptom, COVID Symptom Study, HelpBeatCOVID19, HowWeFeel, and Outbreaks Near Me; see Table 1). In all, 25 programs were excluded as they did not meet the inclusion criteria (Multimedia Appendix 1). Of the original 6 programs, BeatCOVID19Now has since been terminated, reportedly due to control of disease spread in its home location (Australia), resulting in lack of interest in continuing efforts in the United States and other international settings. Hence, BeatCOVID19Now is not included in the tabulated data.
Table 1. Overview of self-reported symptom tracker programs.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Program name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host institution and partners</td>
<td>Carnegie Mellon University Delphi Research Group Facebook COVID Symptom Study^c Outbreaks Near Me</td>
</tr>
<tr>
<td></td>
<td>John Hopkins Bloomberg School of Public Health John Hopkins School of Medicine John Hopkins Whiting School of Engineering John Hopkins Medicine Technology Innovation Center University of Vermont Medical Center Capitol Technology University ITC InforTech Digital Experience</td>
</tr>
<tr>
<td></td>
<td>University of Michigan Kirusa NJ Tech Council NJ Business &amp; Industry Association Walk-In Medical Urgent Care Sills, Cummins &amp; Gross Decagon Strand SpectraMED- DI</td>
</tr>
<tr>
<td></td>
<td>Harvard T.H. Chan School of Public Health Massachusetts General Hospital King's College London Stanford University School of Medicine ZOE COVIDcast^c Outbreaks Near Me Agricultural Health Study^c American Cancer Society^c Aspree XT^c Black Women's Health Study^c California Teachers Study^c Cancer Prevention Study–3^c Dr. Susan Lovec Foundation Gulf Study^c Gutsc The Multiethnic Cohort Study^7 The Sister Study^7 Stand Up to Cancer^c University of Texas School of Public Health</td>
</tr>
<tr>
<td></td>
<td>University of Alabama Alabama Department of Public Health Harvard T.H Chan School of Public Health MIT^d IQSS^e McGovern Institute Howard Hughes Medical Institute Weizmann Institute of Science Pinterest Feeding America Alex's Lemonade Stand Charito Bill &amp; Melinda Gates Foundation Stanford University^c University of Pennsylvania^c</td>
</tr>
<tr>
<td>Location</td>
<td>Pittsburgh, PA Baltimore, MD Ann Arbor, MI Boston, MA Birmingham, AL Boston, MA Boston, MA</td>
</tr>
<tr>
<td>Characteristic</td>
<td>Program name</td>
</tr>
<tr>
<td>---------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>Funding sources</td>
<td>• None</td>
</tr>
<tr>
<td></td>
<td>• White and Case</td>
</tr>
<tr>
<td></td>
<td>• ITC Infotech</td>
</tr>
<tr>
<td></td>
<td>• New Jersey Technology  Control</td>
</tr>
<tr>
<td></td>
<td>• Kirusa</td>
</tr>
<tr>
<td></td>
<td>• Mass General Wellcome Trust (UK)</td>
</tr>
<tr>
<td></td>
<td>• University of Alabama</td>
</tr>
<tr>
<td></td>
<td>• Bill and Melinda Gates Foundation</td>
</tr>
<tr>
<td></td>
<td>• Ending Pandemics</td>
</tr>
<tr>
<td></td>
<td>• Crowdsourcing</td>
</tr>
<tr>
<td></td>
<td>• AtScale</td>
</tr>
<tr>
<td></td>
<td>• AWS</td>
</tr>
<tr>
<td></td>
<td>• BlazeMeter</td>
</tr>
<tr>
<td></td>
<td>• Cloudflare</td>
</tr>
<tr>
<td></td>
<td>• Datalog</td>
</tr>
<tr>
<td></td>
<td>• MongoDB</td>
</tr>
<tr>
<td></td>
<td>• SurveyMonkey</td>
</tr>
<tr>
<td></td>
<td>• TechSoup</td>
</tr>
<tr>
<td></td>
<td>• Bill and Melinda Gates Foundation</td>
</tr>
<tr>
<td></td>
<td>• Crowdsourcing</td>
</tr>
<tr>
<td></td>
<td>• University of Alabama</td>
</tr>
<tr>
<td></td>
<td>• Ending Pandemics</td>
</tr>
<tr>
<td></td>
<td>• Crowdsourcing</td>
</tr>
<tr>
<td></td>
<td>• AtScale</td>
</tr>
<tr>
<td></td>
<td>• AWS</td>
</tr>
<tr>
<td></td>
<td>• BlazeMeter</td>
</tr>
<tr>
<td></td>
<td>• Cloudflare</td>
</tr>
<tr>
<td></td>
<td>• Datalog</td>
</tr>
<tr>
<td></td>
<td>• MongoDB</td>
</tr>
<tr>
<td></td>
<td>• SurveyMonkey</td>
</tr>
<tr>
<td></td>
<td>• TechSoup</td>
</tr>
<tr>
<td>Intended participants (age in years)</td>
<td>• US residents (18+)</td>
</tr>
<tr>
<td></td>
<td>• US residents (13+)</td>
</tr>
<tr>
<td></td>
<td>• Worldwide (18+)</td>
</tr>
<tr>
<td></td>
<td>• US residents (18+) Participants from other internal studies, including RCTs</td>
</tr>
<tr>
<td></td>
<td>• US residents (18+); particular focus on Alabama and neighboring states</td>
</tr>
<tr>
<td></td>
<td>• US residents (18+)</td>
</tr>
<tr>
<td></td>
<td>• US residents (18+)</td>
</tr>
<tr>
<td>Date symptom tracker was initiated</td>
<td>• April 2020</td>
</tr>
<tr>
<td></td>
<td>• April 2020</td>
</tr>
<tr>
<td></td>
<td>• June 2020</td>
</tr>
<tr>
<td></td>
<td>• April 2020</td>
</tr>
<tr>
<td></td>
<td>• Not available</td>
</tr>
<tr>
<td></td>
<td>• April 2020</td>
</tr>
<tr>
<td></td>
<td>• March 2020</td>
</tr>
<tr>
<td>Number of responses to date</td>
<td>• 19,989,000 (+17415000)</td>
</tr>
<tr>
<td></td>
<td>• 215,000</td>
</tr>
<tr>
<td></td>
<td>• 4,651,000 (+4,553,000)</td>
</tr>
<tr>
<td></td>
<td>• 101,000 (+44,000)</td>
</tr>
<tr>
<td></td>
<td>• 5,867,000 (+5,813,000)</td>
</tr>
<tr>
<td>Mechanism of recruiting participants or platform</td>
<td>• Survey via Facebook</td>
</tr>
<tr>
<td></td>
<td>• Apple App Store</td>
</tr>
<tr>
<td></td>
<td>• Google Play Store</td>
</tr>
<tr>
<td></td>
<td>• Apple App Store</td>
</tr>
<tr>
<td></td>
<td>• Google Play Store</td>
</tr>
<tr>
<td></td>
<td>• Apple App Store</td>
</tr>
<tr>
<td></td>
<td>• Google Play Store</td>
</tr>
<tr>
<td></td>
<td>• Web browser</td>
</tr>
<tr>
<td></td>
<td>• Apple App Store</td>
</tr>
<tr>
<td></td>
<td>• Google Play Store</td>
</tr>
<tr>
<td></td>
<td>• Web browser</td>
</tr>
<tr>
<td>Follow-up</td>
<td>• Daily survey prompts via Facebook</td>
</tr>
<tr>
<td></td>
<td>• Daily phone notifications</td>
</tr>
<tr>
<td></td>
<td>• Daily phone notifications</td>
</tr>
<tr>
<td></td>
<td>• Daily phone notifications</td>
</tr>
<tr>
<td></td>
<td>• SMS notifications every three days</td>
</tr>
<tr>
<td></td>
<td>• Customizable phone notifications</td>
</tr>
<tr>
<td></td>
<td>• Daily SMS notifications</td>
</tr>
<tr>
<td>Frequency of reporting</td>
<td>• Daily</td>
</tr>
<tr>
<td></td>
<td>• Daily</td>
</tr>
<tr>
<td></td>
<td>• —</td>
</tr>
<tr>
<td></td>
<td>• Live Data</td>
</tr>
<tr>
<td></td>
<td>• Live Data</td>
</tr>
<tr>
<td></td>
<td>• Daily</td>
</tr>
<tr>
<td></td>
<td>• Weekly</td>
</tr>
<tr>
<td>Availability of summary tables for external synthesis or utilization</td>
<td>• Yes</td>
</tr>
<tr>
<td></td>
<td>• Yes</td>
</tr>
<tr>
<td></td>
<td>• Yes</td>
</tr>
<tr>
<td></td>
<td>• Yes</td>
</tr>
<tr>
<td></td>
<td>• No</td>
</tr>
<tr>
<td></td>
<td>• No</td>
</tr>
<tr>
<td></td>
<td>• Yes</td>
</tr>
<tr>
<td>Intended audience for the product</td>
<td>• Public at large</td>
</tr>
<tr>
<td></td>
<td>• State and local health officials</td>
</tr>
<tr>
<td></td>
<td>• Researchers</td>
</tr>
<tr>
<td></td>
<td>• Public at large</td>
</tr>
<tr>
<td></td>
<td>• Health care providers</td>
</tr>
<tr>
<td></td>
<td>• Researchers</td>
</tr>
<tr>
<td></td>
<td>• Public at large</td>
</tr>
<tr>
<td></td>
<td>• Neighboring States</td>
</tr>
<tr>
<td></td>
<td>• State and local health officials</td>
</tr>
<tr>
<td></td>
<td>• Local policy makers</td>
</tr>
<tr>
<td>Characteristic</td>
<td>Program name</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>COVIDcast</td>
<td>• Public at large</td>
</tr>
<tr>
<td></td>
<td>• State and local public health officials</td>
</tr>
<tr>
<td></td>
<td>• US policy makers</td>
</tr>
<tr>
<td></td>
<td>• Health care providers</td>
</tr>
<tr>
<td></td>
<td>• Health care systems</td>
</tr>
<tr>
<td></td>
<td>• CDC and national public health organizations^c</td>
</tr>
<tr>
<td></td>
<td>• Researchers^c</td>
</tr>
<tr>
<td>COVID Control</td>
<td>• Public at large</td>
</tr>
<tr>
<td></td>
<td>• Participants of internal studies</td>
</tr>
<tr>
<td>COVIDSymptom Study^a</td>
<td>• CDC and national public health organizations^c</td>
</tr>
<tr>
<td></td>
<td>• Health care providers^c</td>
</tr>
<tr>
<td></td>
<td>• Health care systems^c</td>
</tr>
<tr>
<td></td>
<td>• Researchers^c</td>
</tr>
<tr>
<td></td>
<td>• State and local public health officials^c</td>
</tr>
<tr>
<td>HelpBeat-COVID19</td>
<td>• US policy makers^c</td>
</tr>
<tr>
<td>HowWeFeel</td>
<td>• Public at large</td>
</tr>
<tr>
<td></td>
<td>• State and local public health officials</td>
</tr>
<tr>
<td></td>
<td>• Health care providers</td>
</tr>
<tr>
<td></td>
<td>• Health care systems</td>
</tr>
<tr>
<td></td>
<td>• Researchers</td>
</tr>
<tr>
<td></td>
<td>• State and local public health officials</td>
</tr>
<tr>
<td>Outbreaks Near Me^b</td>
<td>• CDC and national public health organizations</td>
</tr>
<tr>
<td></td>
<td>• Researchers</td>
</tr>
<tr>
<td></td>
<td>• Health care providers</td>
</tr>
<tr>
<td></td>
<td>• Health care systems</td>
</tr>
<tr>
<td></td>
<td>• US policy makers</td>
</tr>
<tr>
<td>Publicly available data privacy statement</td>
<td>• Yes</td>
</tr>
</tbody>
</table>

^a Formerly known as COVID Symptom Tracker.
^b Formerly known as COVIDNearYou.
^c Indicates changes made since the previous synthesis.
^d MIT: Massachusetts Institute of Technology.
^e IQSS: Institute for Quantitative Social Science.
^f AWS: Amazon Web Services.
^g RCT: randomized controlled trial.
^h Not available.
^i CDC: Centers of Disease Control and Prevention.

All 6 of the previously reported programs have maintained their affiliation with a university, school of medicine, or school of public health. The 2 newly added programs are also affiliated with universities, further emphasizing the importance of academic institutions in studying the spread of COVID-19 in the United States. Three programs continue to be based in Boston, MA (ie, COVID Symptom Tracker, HowWeFeel, and Outbreaks Near Me), with another affiliated through partnerships (COVIDcast). However, the 2 new programs are based in Ann Arbor, MI and Baltimore, MD.

All of the programs continued to receive responses through electronic reporting mechanisms, with 2 programs utilizing a web browser-based approach (ie, Outbreaks Near Me and HelpBeatCOVID19). Four programs, including both the newly added programs, use apps for both Apple and Android devices; these include COVID Control, COVIDSymptom, COVID Symptom Study, and HowWeFeel. One program (COVIDcast) uses social media as a platform for their survey.

Upon reviewing the data elements being collected across the programs (Table 2), it appears that some of the trackers have been revised to ask more timely questions, including those covering topics like vaccination (eg, COVIDcast, COVID Symptom Study, HowWeFeel, Outbreaks Near Me), vaccine hesitancy (eg, COVIDcast and HowWeFeel), adherence to social distancing recommendations put forth by the Centers for Disease Control and Prevention (CDC; eg, COVIDcast and HowWeFeel), and in-person schooling for children K-12 (eg, HowWeFeel). Notably, COVIDcast asks about mask usage. The remaining programs (COVID Control, COVIDSymptom, HelpBeatCOVID19, and Outbreaks Near Me) continue to focus their efforts only on elements such as ongoing symptoms, testing status, and demographic information. Interestingly, COVIDcast and COVID Symptom Tracker have limited or removed much of the demographic data that they collect.
Table 2. Data elements across programs.

<table>
<thead>
<tr>
<th>Question</th>
<th>COVIDcast Control</th>
<th>COVID Symptom Study</th>
<th>HelpBeatCOVID19</th>
<th>HowWeFeel</th>
<th>Outbreaks Near Me</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are you completing the survey on behalf of someone else?</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Age</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Gender</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Zip code</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Race or ethnicity</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Blood group</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Symptoms at time of check-in</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Symptoms within the last 24 hours</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Symptoms over the past 7 days</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Date of symptom onset</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Do others in your household have similar symptoms to those you reported?</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Temperature at the time of check-in</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Highest temperature over symptom duration</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Hours of sleep previous night</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Have you been tested for COVID-19?</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Type of testing or care sought due to symptoms</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>What type of medical test did you receive</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Results of testing</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>In the past 14 days, did you want a COVID-19 test but did not receive one</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>How long after you started feeling ill did you see a health professional?</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>What prescription, if any, did you receive for your illness?</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Have you received the COVID-19 vaccine?</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Would you accept a COVID-19 vaccine if offered?</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>If a safe, effective coronavirus vaccine were available, how likely would you be to get vaccinated?</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>What is the main reason you got the COVID-19 vaccine?</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>What is the main reason you did not get the COVID-19 vaccine?</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Do you know anyone who has received a COVID-19 vaccine?</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>If you have a child under 18, how likely are you to get your child vaccinated?</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Are you experiencing any symptoms near the injection site?</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Preexisting conditions</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Obesity</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Are you pregnant?</td>
<td>✓✓✓✓✓</td>
<td>✓✓✓✓✓✓</td>
<td>✓✓✓✓</td>
<td>✓✓</td>
<td>✓✓✓✓✓✓</td>
</tr>
<tr>
<td>Question</td>
<td>COVIDcast</td>
<td>COVID Control</td>
<td>COVID Symptom Study</td>
<td>HelpBeatCOVID19</td>
<td>HowWeFeel</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>-----------</td>
<td>---------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>------------</td>
</tr>
<tr>
<td>Are you/have you ever been a smoker</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Have you received the flu vaccine?</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Number of people in the household</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are you a parent?</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of children in the household</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do any children in your household (pre-K through grade 12) go to full-time in-person classes?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has anyone in your household been diagnosed with COVID-19?</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of domicile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Have you left your home in the past 24 hours?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Reason for leaving home</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What protective measures did you take when you left home?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>In the past 7 days, did you wear a mask most or all of the time in public?</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In the past 7 days, when you were in public places where social distancing is not possible, did most or all other people wear masks?</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In the past 24 hours, did you spend time indoors with someone who isn’t currently staying with you?</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In the past 24 hours, did you attend an indoor event with more than 10 people?</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In the past 24 hours, did you go to an indoor market, grocery store, or pharmacy?</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In the past 24 hours, did you have a meal or drink indoors at a bar, restaurant, or cafe?</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In the past 24 hours, did you use public transit?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In the past 7 days, have you traveled outside of your state?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Have you been in contact with anyone diagnosed with COVID-19?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Do you personally know someone in your local community who has COVID-like symptoms?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does your job require you to leave your home and go to another place to work where you come in contact with public?</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How much of the following feelings (tired, calm, happy, angry, sad, thoughtful, optimistic, anxious, lonely, grateful, hopeful, stressed) have you felt so far today?</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do you feel very or somewhat worried about becoming seriously ill from COVID-19?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>
All 7 currently active programs collect information regarding symptoms, although the time period varies, with the majority (4/7, 57%), asking participants to report current symptoms at the time of check-in (ie, COVID Control, COVID Symptom Tracker, and HelpBeatCOVID19), 2 (29%) asking about symptoms experienced over the past week (ie, HelpBeatCOVID19 and Outbreaks Near Me), and only 1 (14%) asking about symptoms experienced over the past 24 hours (ie, COVIDcast). Except for COVIDcast, all other programs ask symptomatic participants who report fever for their temperature, with COVID Control uniquely asking participants who report to have experienced no symptoms. All of these programs (excluding COVIDcast) also ask participants whether they have undergone testing, and 5 programs (ie, COVID Control, COVID Symptom Tracker, COVIDcast, HelpBeatCOVID19, Outbreaks Near Me) ask about the results. COVID Control and COVID Symptom Tracker limited their questions to symptoms, testing, and demographic data, whereas other programs had unique questions, indicating their areas of focus. COVIDcast tailored its questions to current issues, including vaccination status, adherence to social distancing guidelines, mask usage, use of public transit, travel, and mental health concerns. COVID Symptom Study also asked participants about vaccination status, as well as vaccine hesitancy, vaccine side effects, and it was the only program to ask participants about their blood group. HelpBeatCOVID19 asked participants about their household, health insurance status, and financial insecurity. HelpBeatCOVID19 asked questions on vaccination status; vaccine hesitancy, including asking participants why they have or have not been vaccinated and whether they would vaccinate their children; adherence to social distancing guidelines; and mental health concerns. Finally, Outbreaks Near Me asks participants in-depth questions about testing, including what kind of test participants receive, their diagnosis, and any medications they are prescribed.

## Discussion

### Principal Findings

Over the course of the past year, symptom tracking programs have been exceedingly useful for predictive modeling and population research throughout the COVID-19 pandemic [10]. The data from COVID Symptom Study, for example, has already been used to unearth COVID-19–related trends, even revealing 6 distinct “types” of COVID-19 emerging among their participants. The importance of their contributions were highlighted in the *New England Journal of Medicine* article “Putting the Public Back in Public Health—Surveying Symptoms of Covid-19” [10]. This framework update shows the increase in responses, collaboration, and evolution of questionnaires between programs exemplifying their use by the public health community.

Many symptom checking programs (not included in this framework), defined—for the purposes of this study—as programs that serve to indicate whether the participant likely should or should not seek medical attention without aiming to record symptoms over time, have emerged and with good reason [11]. These programs are critical in deterring those without serious symptoms from seeking emergency care. However, the number of symptom tracking programs has not correspondingly increased. Only 2 additional symptom tracking programs were identified, which indicates either the demand for these programs is significantly lower than symptom checking programs or that the current programs are doing well to serve this niche. It is also possible that both reasons are true.

Reported responses to each of the original 6 programs substantially increased. The lowest number of responses, defined as unique symptom entries by an individual, were received for HelpBeatCOVID19—recorded at 101,000, an increase of 44,000. The remaining 3 programs with reported responses also showed exponential increase in the number of responses. For instance, COVIDcast had 19,989,000 responses, an increase of 17,415,000—the highest of all programs. COVID Symptom Study recorded 4,651,000 responses, an increase by 4,553,000. Outbreaks Near Me recorded 5,867,000 responses, an increase of 5,813,000. In the original framework, two-thirds of the programs had less than 100,000 responses. Among the newly added programs, COVID Control had 215,000 responses. Although only half of the previous programs focused on user retention through notifications (ie, COVID Symptom Study, Outbreaks Near Me, and HelpBeatCOVID19), we found that all programs now prompted previous participants on a daily or semiregular basis through either phone or social media notifications or SMS text messages.

Despite the addition of 2 new programs to the framework, 1 program concluded during the past year. BeatCOVID19Now, based in Australia, concluded and terminated its website in early 2021. Notably, it was the only program in the original synthesis that served international participants and was founded by the same researchers who developed the Influenza Intensity and

---

**Table 1: Questions asked by each program**

<table>
<thead>
<tr>
<th>Question</th>
<th>COVIDcast</th>
<th>COVID Control</th>
<th>COVID Symptom Tracker</th>
<th>COVID Symptom Study</th>
<th>HelpBeatCOVID19</th>
<th>HowWeFeel</th>
<th>Outbreaks Near Me</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you feel very or somewhat worried about your household’s finances for the next month?</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health insurance status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If you needed to get to a hospital or testing center, how would you get there?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Can you afford any payment or co-payment required for services?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

---

https://formative.jmir.org/2021/12/e31271

**JMIR Form Res 2021 | vol. 5 | iss. 12 | e31271 | p.283**

*(page number not for citation purposes)*
Impact Questionnaire (fluiQ) [12]. The discontinuation of BeatCOVID19Now also raises the question of the longevity of the other programs—like BeatCOVID19Now, Outbreaks Near Me was developed based on existing infrastructure, namely Flu Near You, which performed a similar function for yearly influenza outbreaks. Therefore, discontinuation of BeatCOVID19Now poses the question of whether other programs may be discontinued in the near future.

Collaboration between programs also seems to be auspicious for growth and user retention. Three of the programs (ie, COVIdcast, COVID Symptom Tracker, and Outbreaks Near Me) have begun collaborating. Lack of collaboration was a chief concern in the first iteration of our framework synthesis, with the concern being that it would lead to unnecessary duplication and a division of the pool of potential participants, and ultimately, all of the programs sought to share this important data with those who could best utilize it. Although these concerns cannot be completely assuaged without further collaboration among groups, initiating collaboration among the 3 programs with the highest response rates is a step forward.

A notable development was the continued change in data elements observed by the majority of programs. Elements such as symptom status and potential contact continue to be necessary; however, seeing programs adapt to changes in pandemic response by adding elements concerning vaccination, vaccine hesitancy, social distancing, and mental health indicates what could be the next direction for these programs.

Limitations
As with the first iteration of this study, several limitations must be acknowledged. First, our analysis was limited to English-language programs; therefore, we may have missed nuances of data collection that are more important to non–English-speaking residents. Second, although the speed of framework analysis enables rapid evaluation of commonalities, it does not provide the in-depth rigor of a full systematic review.

Third, our collected data evaluated differences in the number of responses to each program, but it only pertained to what data was collected, how that data is being collected, and what groups created and funded each program. Analyzing the effectiveness, market penetration, or user demographics of evaluated programs is beyond the scope of this study. This kind of analysis, if undertaken in a future study, may provide insight into the question of potential longevity posed by this framework. Fourth, we recognize that program participation is limited to only those who have access to the internet or cellular phone service, creating an unintended disparity among respondents based on their access to and utilization of technology. Therefore, the underlying reasons for the difference in response rate remain beyond the scope of this study. Last, this synthesis neither provides critical appraisal of programs nor evaluates programs for effectiveness.

Conclusions
At this time, the future of self-reported symptom tracking for COVID-19 and for these programs, is unclear. Some sources have speculated that COVID-19 may become a yearly occurrence much like the flu, and if so, the data that these programs generate is still valuable [13]. However, it is unclear whether the public will maintain the same level of interest in reporting their symptoms on a regular basis if the prevalence of COVID-19 becomes more common, and, for programs like COVIdcast, whether the platform where the survey is hosted will continue to support them. The aforementioned conclusion of BeatCOVID19Now represents a foreboding that could be lying ahead—that lack of interest after virus control through the roll-out of rapid and widely available testing and successful vaccination could lead to a lack of interest among participants, and ultimately discontinuation; however, as of May 2021, these 7 programs have not shown any signs of waning.

Acknowledgments
The contents of this publication are the sole responsibility of the authors and do not necessarily reflect the views, assertions, opinions or policies of the Uniformed Services University of the Health Sciences (USUHS); the Henry M Jackson Foundation for the Advancement of Military Medicine, Inc (HJF); the Department of Defense (DoD); or the Departments of the Army, Navy, or Air Force. Mention of trade names, commercial products, or organizations does not imply endorsement by the US Government.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Programs excluded from the study.

References


Abbreviations

DoD: Department of Defense

fluiiQ: Influenza Intensity and Impact Questionnaire

HJF: Henry M Jackson Foundation for the Advancement of Military Medicine, Inc

USUH: Uniformed Services University of the Health Sciences

©Miranda Lynn Janvrin, Jessica Korona-Bailey, Tracey Pérez Koehlmoos. Originally published in JMIR Formative Research (https://formative.jmir.org), 06.12.2021. 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.
Abstract

**Background:** Social media has emerged as an effective means of information sharing and community building among health professionals. The utility of these platforms is likely heightened during times of health system crises and global uncertainty. Studies have demonstrated that physicians’ social media platforms serve to bridge the gap of information between on-the-ground experiences of health care workers and emerging knowledge.

**Objective:** The primary aim of this study was to characterize the use of a physician WhatsApp (WhatsApp LLC) group chat during the early months of the COVID-19 pandemic.

**Methods:** Through the lens of the social network theory, we performed a qualitative content analysis of the posts of a women physician WhatsApp group located in the United Arab Emirates between February 1, 2020, and May 31, 2020, that is, during the initial surge of COVID-19 cases.

**Results:** There were 6101 posts during the study period, which reflected a 2.6-fold increase in platform use when compared with platform use in the year prior. A total of 8 themes and 9 subthemes were described. The top 3 uses of the platform were requests for information (posts: 2818/6101, 46.2%), member support and promotion (posts: 988/6101, 16.2%), and information sharing (posts: 896/6101, 14.7%). A substantial proportion of posts were related to COVID-19 (2653/6101, 43.5%), with the most popular theme being requests for logistical (nonmedical) information. Among posts containing COVID-19–related medical information, it was notable that two-thirds (571/868, 65.8%) of these posts were from public mass media or unverified sources.

**Conclusions:** Health crises can potentiate the use of social media platforms among physicians. This reflects physicians’ tendency to turn to these platforms for information sharing and community building purposes. However, important questions remain regarding the accuracy and credibility of the information shared. Our findings suggest that the training of physicians in social media practices and information dissemination may be needed.

**(JMIR Form Res 2021;5(12):e31791)** doi:10.2196/31791

**KEYWORDS**

WhatsApp; social media; physician; pandemic; COVID-19; qualitative; communication; misinformation; information-seeking behavior; information seeking; information sharing; content analysis; community
including curfews, lockdowns, social distancing restrictions, and testing and tracing requirements, presents additional challenges. In order to ensure their personal safety and provide care, frontline medical professionals need to be equipped with the most evidence-based clinical pathways and public health protocols.

**Use of Social Media**

Several studies have shown that physicians’ social media platforms have become significant facilitators of bridging gaps of information between on-the-ground experiences of health care workers and emerging, scientific, clinical, and population-level knowledge [2,3]. Researchers have analyzed various social media platforms to better understand their use in public health discourse [4,5]. Moreover, infoveillance studies have confirmed a marked increase in individuals’ activity on social media platforms, particularly during the COVID-19 pandemic [6]. For example, a COVID-19 physician group, which was created on Facebook in March 2020 and was described as “an inclusive resource for physicians to share frontline clinical information about COVID-19 as it becomes available,” quickly rose to considerable popularity; the group has approximately 29,000 members to date [7]. Other studies have described how health care providers worldwide have used Twitter and WhatsApp (WhatsApp LLC) during the pandemic to disseminate news and discoveries to colleagues and communicate health information directly to patients [2,8]. There is limited published information however on the use of social media among groups of physicians during a medical and public health crisis.

We previously reported an analysis of the WhatsApp posts of a women physician group; we noted that the platform was effective in enabling female physicians to expand networks, exchange ideas, share scientific information, celebrate accomplishments, and provide support to colleagues [9]. In this study, by using social network theory as an overarching lens [10], we sought to analyze the content of the social media interactions of this group’s members during a public health crisis. The primary purpose of this study was to characterize the use of a social media platform among members of a physician group during the COVID-19 pandemic. Our additional aims included examining group members’ levels of engagement (ie, by comparing them to members’ pre-pandemic levels of engagement) and identifying the sources of medical information that were shared among the physician members.

**Methods**

**Setting and Population**

On January 23, 2020, the United Arab Emirates (UAE) reported its first confirmed case of COVID-19. Over the following 4 months, the country experienced a surge of cases; over 60,000 UAE patients were infected with SARS-CoV-2 [11]. Prior to data extraction and analysis, group members were informed about this retrospective study via a WhatsApp message and were given the opportunity to have their posts excluded from analysis. Data were exported from the WhatsApp Messenger group to Microsoft Office Excel 2013 by one of the researchers (PA), who removed identifying information; retained the content, dates, and times of posts; and assigned each member a unique numeric identifier to calculate the percentage of members who wrote posts. All data were then anonymized for qualitative analysis. The messages were analyzed via qualitative content analysis [12]. We approached the data through the lens of the social network theory, which focuses on the effect of social relationships on processing media influence, transferring information, and enabling attitudinal or behavioral change [10]. Two physician researchers (HI and PA) independently coded each post, performed a content analysis of the messages, and produced a list of the common themes that they identified. After this initial review, the researchers discussed their findings, and through discussion, they reached consensus on the themes and created subthemes. The two primary reviewers then independently categorized all posts according to the predetermined themes. Any disagreements were resolved by consensus, and any remaining discordance was brought to the third physician researcher (SAR) and discussed until consensus was achieved. A descriptive quantitative analysis was conducted by using Microsoft Excel 2019 to analyze the frequency of posts within each identified theme. This study was reviewed by the Cleveland Clinic Abu Dhabi Research Ethics Committee and was deemed exempt from institutional review board review, as the data were retrospective, were deidentified, and did not involve any patient information.

**Team Reflexivity**

We were cognizant that our research team consisted of 3 female physicians who lived and worked in the Middle East and were frontline workers during the COVID-19 pandemic. To minimize bias, we were blind to participants’ identities. We were mindful of how our experiences influenced our analysis of the data and engaged in frequent group conversations to share, support, and challenge each other’s interpretations.

**Results**

At the time of data extraction and analysis, there were 161 group members—a 32% increase from the 122 members in the prior year. From February 1, 2020, to May 31, 2020, there were a total of 6101 posts. Of the 161 members, 130 (80.7%) posted at least once during this time period. The number of posts increased 2.6-fold from the number of posts during the same
time period in 2019. Further, 1204 more messages were posted in the chat during this 5-month study interval than during the entire preceding year (6101 posts vs 4897 posts, respectively). Approximately half of all posts (2653/6101, 43.5%) were directly related to the COVID-19 pandemic.

There were 8 general themes identified. Table 1 provides a description of themes and subthemes, example posts, and the relative frequencies of each theme. The most frequent theme was related to requests for information, which represented 46.2% (2818/6101) of all posts. Of the information requests, the majority (1308/2818, 46.4%) were requests for non–COVID-19–related general medical information, and many of these posts consisted of physician referral requests. Of the COVID-19–related information requests, the vast majority (945/1184, 79.8%) were related to logistical information, including quarantine measures, school closures, or mask mandates. Only 3.9% (239/6101) of all posts consisted of specific diagnostic or treatment queries regarding COVID-19.

Approximately 15% (896/6101, 14.7%) of the posts consisted of medical information that was shared with the group by individual members. There were twice as many posts containing information from unverified and non–evidence-based sources (n=517; eg, blogs, social media messages, and local newspaper articles) as there were posts containing evidence-based information (n=297). Several group members expressed confusion and frustration. One group participant noted the following:

I have a headache from all the COVID-19 stuff I am reading. I no longer know what to trust and who to believe.

Another physician stated:

This article a perfect example about why a lot of the social media posting of drafts, small series, unproven theories, personal opinion etc. is frankly dangerous. We should defer to only published peer reviewed papers and guidelines. I am personally overwhelmed with all the misinformation I get.

The frequencies of posts related to each COVID-19 subtheme are displayed in Table 2. In total, 35.6% (945/2653) of COVID-19–related posts were requests for logistical information, and 22.7% (601/2653) of such posts contained supportive or promotional messages related to COVID-19.
Table 1. Themes derived from the qualitative analysis of the WhatsApp group chat (posts: N=6101).

<table>
<thead>
<tr>
<th>Social network theory principles, general themes, and subthemes</th>
<th>Description of theme</th>
<th>Posts, n (%)</th>
<th>Example posts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relationships in the context of general information seeking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Request for information</strong></td>
<td>Member requesting information from other group members</td>
<td>2818 (46.2)</td>
<td>• N/A</td>
</tr>
<tr>
<td>General (medical)</td>
<td>Information sought on any non–COVID-19–related medical</td>
<td>1308 (21.4)</td>
<td>• “I need a recommendation for a fertility/IVF center in Abu Dhabi.”</td>
</tr>
<tr>
<td></td>
<td>matter</td>
<td></td>
<td>• “Any pediatric urologist in the group, or one that can be recommended?”</td>
</tr>
<tr>
<td>COVID-19 logistics</td>
<td>Information sought on logistical topics, such as the</td>
<td>945 (15.5)</td>
<td>• “Question for the OB in the group: What are your PPE protocol in labor and delivery, patient of unknown COVID status? Any guidelines or protocols on that?”</td>
</tr>
<tr>
<td></td>
<td>location of testing sites, curfew rules, and personal</td>
<td></td>
<td>• “Does this mean no school for 4 weeks starting this Sunday?”</td>
</tr>
<tr>
<td></td>
<td>protective equipment protocols</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General (nonmedical)</td>
<td>Information sought on any nonmedical matter</td>
<td>326 (5.3)</td>
<td>• “Dear ladies, any houseplant experts in the group, particularly orchids.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• “Any recommendations for piano repair who will visit your home?”</td>
</tr>
<tr>
<td>COVID-19 medical</td>
<td>Information sought on COVID-19 diagnoses, symptoms, and</td>
<td>239 (3.9)</td>
<td>• “Good evening ladies Is there any evidence that fasting and or associated dehydration is a risk favor for worse outcomes in covid19? Just wondering as Ramadan is almost upon us.”</td>
</tr>
<tr>
<td></td>
<td>treatment protocols</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relationships in the context of community building</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Support and promotion</strong></td>
<td>Member providing moral or emotional support to other</td>
<td>988 (16.2)</td>
<td>• N/A</td>
</tr>
<tr>
<td>COVID-19 related</td>
<td>group members</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>Support given for any non–COVID-19–related matters</td>
<td>387 (6.3)</td>
<td>• “Watch interview with our very own member educating the public on the crucial role anesthesiologists play in delivering safe patient care.”</td>
</tr>
<tr>
<td>Community engagement</td>
<td>Non–medical-related general posts, quotes, memes, videos,</td>
<td>637 (10.4)</td>
<td>• “I second your gratitude for all the blessings we have...and most importantly each other. Once this is over, I plan to make hugs mandatory amongst all</td>
</tr>
<tr>
<td></td>
<td>or articles of interest</td>
<td></td>
<td>Wonder members”</td>
</tr>
<tr>
<td>Celebration</td>
<td>Secular and religious holiday wishes, personal and</td>
<td>350 (5.7)</td>
<td>• “Beautiful baby! Wishing him a life full of happiness, health and prosperity.”</td>
</tr>
<tr>
<td></td>
<td>professional milestones, and celebrations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group administration</td>
<td>Posts for announcing group activities and events as well</td>
<td>250 (4.1)</td>
<td>• “Reminder: Our Zoom meet is tonight at 7:30p. We have a dozen WONDER docs signed up already. Anyone else wants to join? DM me your email address.</td>
</tr>
<tr>
<td></td>
<td>as for adding and welcoming new group members</td>
<td></td>
<td>Looking forward to catching up!”</td>
</tr>
<tr>
<td>Women empowerment</td>
<td>Articles regarding women in medicine, inspirational</td>
<td>73 (1.2)</td>
<td>• “Meet the top 10 Power Businesswomen in the Middle East’ ranked by Forbes”</td>
</tr>
<tr>
<td></td>
<td>quotes, and images related to women’s empowerment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relationships in the context of information sharing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Information sharing</strong></td>
<td>Member sharing information with the group</td>
<td>896 (14.7)</td>
<td>• N/A</td>
</tr>
</tbody>
</table>
Table 2. Analysis of the subset of posts related to COVID-19 (n=2653).

<table>
<thead>
<tr>
<th>Subtheme</th>
<th>COVID-19–related posts, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requests for logistical information related to COVID-19</td>
<td>945 (35.6)</td>
</tr>
<tr>
<td>Requests for medical information related to COVID-19</td>
<td>239 (9)</td>
</tr>
<tr>
<td>Sharing of COVID-19–related information from peer-reviewed literature and verifiable sources</td>
<td>297 (11.2)</td>
</tr>
<tr>
<td>Sharing of COVID-19–related information from social media and unverifiable sources</td>
<td>571 (21.5)</td>
</tr>
<tr>
<td>Supportive, encouraging, or promotional messages related to COVID-19</td>
<td>601 (22.7)</td>
</tr>
</tbody>
</table>

Discussion

Use of Social Media

Social media content can provide important insights into the issues and topics that concern health care providers during a global health crisis. Our report on the use of a physician group’s WhatsApp chat during the evolving COVID-19 pandemic demonstrates the increased use of this social media forum, and a substantial proportion of the content was related to COVID-19. Social network theory emphasizes the importance of relationships in the context of general information seeking, knowledge sharing, and community building [10]. Accordingly, our study identified 3 main aims of the use of the WhatsApp group, namely information requests, information dissemination, and support and encouragement. A major premise of the social network theory is that a wide network of weaker relationships allows for access to more individuals and resources and can therefore be more beneficial than 1 or 2 strong ties [10]. Throughout the early months of 2020, health care workers cared for large volumes of critically ill patients without any evidence-based therapies while simultaneously dealing with an onslaught of research findings and information [1]. Our findings reflect the confusion and frustration that are often felt by frontline physicians who are trying to navigate a global public health emergency for themselves, their families, and their patients. In our study, group chat members often turned to their colleagues for advice and support. This was evidenced by the large volume of conversations that occurred during this time period and the high engagement levels of members. Compared to the number of posts from the same time period in the previous year, the number of posts increased substantially in 2020. Although there was a modest increase in membership, the exponential rise in the number of posts likely represents the increased use of the platform as a resource for physicians during the extraordinary circumstance of a global pandemic. In fact, 1204 more posts were found during our 5-month study interval than during the entire preceding year (6101 posts vs 4897 posts, respectively). Other studies have reported increased social media use during the COVID-19 pandemic [2,3,6,9].

Almost half (2653/6101, 43.5%) of the posts were directly related to COVID-19, and over one-third (945/2653, 35.6%) of these posts consisted of requests for logistical information. The non–COVID-19–related posts highlighted several important points. For instance, even in the midst of a global pandemic, physicians continued to provide general medical care to their patient populations. Additionally, group members frequently discussed non–COVID-19–related and nonmedical interests. As such, the chat group likely allowed for the opportunity to provide a sense of normalcy to frontline workers in the midst of a public health crisis. It is interesting that the vast majority
of information requests related to COVID-19 dealt with logistics rather than with medical or treatment queries (945/1184, 79.8%). This may have been due to the lack of information available at the time about the novel SARS-CoV-2, but this may also signify the general confusion on public health protocols, which grew as lockdowns, school closures, and social distancing restrictions were implemented. In addition to personally navigating these regulations, physicians were required to guide patients through testing, isolation, and quarantine procedures, often with limited medical knowledge and under frequently changing government policies. Further, as the studied posts were from a women physician group, it should be noted that the group members likely had primary caregiver roles within their families and were personally impacted by the COVID-19 public health protocols. Studies have confirmed that women physicians often bear the majority of childcare and household responsibilities [13,14]. The focus of posts on COVID-19–related logistics may therefore reflect day-to-day priorities that may be shared among female group members. It is notable that community engagement, celebration, and promotion collectively remained important themes during the pandemic. This reflects the tendency of group members to encourage and support each other during uncertain times, which further reinforces the critical role of social media platforms in facilitating a sense of community among medical professionals.

Study Implications
We are concerned that a substantial majority (571/868, 65.8%) of the posts in the WhatsApp group that contained COVID-19–related information often cited media, social media, or unnamed sources. In fact, only one-third (297/868, 34.2%) of such posts contained information from published medical literature or other verifiable sources. The sharing and discussion of various medical messages within a physician group could serve as a means of critiquing the veracity of information or creating awareness of web-based misinformation. However, the sheer volume of unverified posts and the frequent deviations from the evidence-based data that physicians are expected to disseminate could result in confusion and the inadvertent spread of misinformation from physicians to their patients, as evidenced by several chat participants who expressed concern and frustration about the posting of unverified information by other group members. This phenomenon is not unique to this one chat group. A previous study on the prevalence of misinformation in tweets about health care found that approximately 20% of tweets were inaccurate [15]. In fact, one of the most widely spread conspiracy theories, which linked COVID-19 to the 5G network, was traced to comments made by a Belgian physician in January 2020 [16]. The director-general of the World Health Organization described the large volume of unproven or inaccurate information on social media during the pandemic as an infodemic of misinformation [17]. The confusion resulting from misinformation can ultimately cause physicians to question the legitimacy of new scientific discoveries regarding effective COVID-19 therapeutics or vaccines. Moreover, the high volume of social media posts can result in information fatigue. Despite the understandable desire of physicians to share and receive useful information, without verifying the veracity and credibility of information prior to sharing it, physicians risk unwittingly facilitating the dissemination of misinformation. This may reflect a lack of formal social media education among physicians [18], which is compounded by the considerable challenges of analyzing data during an evolving pandemic [19]. More and more social interactions are occurring on web-based platforms. As such, physicians should be more cognizant of appropriate and effective social media use in the context of data analysis, synthesis, and sharing. Our findings have several important implications. First, the results substantiate the critical role that social media platforms play in facilitating communication and fostering connectedness among physicians coping with population health crises. Second, our study also provides insights on content and topics that seem to be the most relevant to physician communities during such crises. Lastly, our data reflect areas of concern regarding the use of social media in these professional communities during times of uncertainty and can be used to inform the design of future interventions and research.

In addition to confirming the results of prior work, this study highlights the need for additional research into the evidence-based approaches that physicians use to analyze health information obtained from social media. Similar concerns have been raised about the increasing dissemination of medical information through social media, including the lack of editorial oversight for web-based data and the harm caused by the rapid dissemination of incorrect medical information [20].

Limitations
Although this study involves a single international physician group, the findings likely reflect the common challenges faced by health care workers who deliver health services in dynamic logistical and biomedical environments that are intensified by global health crises. The study group contained only women physicians; therefore, our findings may represent priorities that correlate with gendered experiences and roles. To our knowledge, the existing literature lacks published studies that report on gender differences in physicians’ social media behaviors, though studies have suggested that there are gender differences in success on social media platforms [21,22]. Lastly, it is unknown whether group messages actually impacted physicians’ attitudes or behaviors.

Conclusion
The uncertainties posed by an evolving global health crisis represent considerable challenges to the health care workforce. As the world has become increasingly more connected through social media, these platforms represent critical information dissemination tools. Our findings confirmed the importance of social media in creating a communicative and collaborative platform for physicians in the midst of a public health emergency. Although more accessible information can undoubtedly benefit patient care, our findings raised important questions regarding the accuracy and credibility of shared information. Larger multinational infoveillance studies are needed to better understand social media discourse among physicians during public health crises.
Authors’ Contributions
SAR led the drafting of the manuscript. All authors contributed equally to the conception, analysis, editing, and final approval of the manuscript. All authors had equal access to the data and had final responsibility for the publication of this study.

Conflicts of Interest
None declared.

References
7. COVID-19 USA Physician/APP group. Facebook. URL: https://www.facebook.com/groups/815998962228639/ [accessed 2020-12-30]


Abbreviations

UAE: United Arab Emirates

WONDER: Women Doctors in the Emirates