# Contents

## Original Papers

**A Smartphone App to Support Adherence to Inhaled Corticosteroids in Young Adults With Asthma: Multi-Methods Feasibility Study (e28784)**

Jane Murphy, Jenny McSharry, Lisa Hynes, Gerard Molloy .......................................................... 3

**Ecological Momentary Assessment of Bipolar Disorder Symptoms and Partner Affect: Longitudinal Pilot Study (e30472)**

Mor Yerushalmi, Andrew Sixsmith, Ariel Pollock Star, David King, Norm O’Rourke. .......................... 23

**Recovery Following Peer and Text Messaging Support After Discharge From Acute Psychiatric Care in Edmonton, Alberta: Controlled Observational Study (e27137)**

Reham Shalaby, Marianne Hrabok, Pamela Spurvy, Rabab Abou El-Magd, Michelle Knox, Rebecca Rude, Wesley Vuong, Shireen Surood, Liana Uruchuk, Mark Snaterse, Andrew Greenshaw, Xin-Min Li, Vincent Agyapong .......................... 33

**Investigation of the Effects of an Online Support Group for Mental Health Problems on Stigma and Help-Seeking Among Japanese Adults: Cross-sectional Study (e21348)**

Osamu Kobori, Naoki Yoshinaga .......................................................... 45

**Using the Think-Aloud Method to Assess the Feasibility and Acceptability of Network Canvas Among Black Men Who Have Sex With Men and Transgender Persons: Qualitative Analysis (e30237)**

Natalie Crawford, Donie Josma, Kristin Harrington, Joseph Morris, Alvan Quamina, Michelle Birkett, Gregory Phillips II .......................................................... 63

**Collecting Social Media Information in a Substance Use Intervention Trial With Adolescent Girls With Lifetime Substance Use History: Observational Study (e25405)**

Lili Ramos, Joseline Delgadillo, Sarah Velez, Emily Dauria, Jamie Salas, Marina Tolou-Shams. 71

**A Suite of Mobile Conversational Agents for Daily Stress Management (Popbots): Mixed Methods Exploratory Study (e25294)**

Matthew Mauriello, Nantanick Tantivasadakarn, Marco Mora-Mendoza, Emmanuel Lincoln, Grace Hon, Parsa Nowruzi, Doreen Simon, Luke Hansen, Nathaniel Goenawan, Joshua Kim, Nikhil Gowda, Dan Jurafsky, Pablo Paredes. .......................................................... 77

**Semisupervised Deep Learning Techniques for Predicting Acute Respiratory Distress Syndrome From Time-Series Clinical Data: Model Development and Validation Study (e28028)**

Carson Lam, Chak Tso, Abigail Green-Saxena, Emily Pellegrini, Zohora Iqbal, Daniel Evans, Jana Hoffman, Jacob Calvert, Qingqing Mao, Ritankar Das .................. 93
Assessing the Care Modality Preferences and Predictors for Digital Mental Health Treatment Seekers in a Technology-Enabled Stepped Care Delivery System: Cross-sectional Study (e30162)
Elissa Kozlov, Meghan McDarby, Maximo Prescott, Myra Altman. ................................................................. 105

Designing User-Centered Mobile Health Initiatives to Promote Healthy Behaviors for Children With Disabilities: Development and Usability Study (e23877)

Electronic Video Consent to Power Precision Research: A Pilot Cohort Study (e29123)
Arash Naeim, Sarah Dry, David Elashoff, Zhuoer Xie, Antonia Petrusse, Clara Magyar, Lillian Johansen, Gabriela Werre, Clara Lajonchere, Neil Wenger. ................................................................. 132

Social Networking Site Use During the COVID-19 Pandemic and Its Associations With Social and Emotional Well-being in College Students: Survey Study (e26513)
Alison Tuck, Renee Thompson. ............................................................................................................................. 143

Early Detection of Symptom Exacerbation in Patients With SARS-CoV-2 Infection Using the Fitbit Charge 3 (DEXTERITY): Pilot Evaluation (e30819)
Kan Yamagami, Akihiro Nomura, Mitsuhiro Kometani, Masaya Shimojima, Kenji Sakata, Soichiro Usui, Kenji Furukawa, Masayuki Takamura, Masaki Okajima, Kazuyoshi Watanabe, Takashi Yoneda. ................................................................. 165

Short Paper
Technology Acceptance Among Patients With Hemophilia in Hong Kong and Their Expectations of a Mobile Health App to Promote Self-management: Survey Study (e27985)
Yin Cheung, Pok Lam, Teddy Lam, Henry Lam, Chi Li. .......................................................................................... 55

Viewpoint
Domestic Violence and Mental Health During the COVID-19 Pandemic in Bangladesh (e24624)
Tanjir Rashid Soron, Md Ashiq, Marzia Al-Hakeem, Zaid Chowdhury, Helai Uddin Ahmed, Chaman Afroz Chowdhury. ................................................................. 159
A Smartphone App to Support Adherence to Inhaled Corticosteroids in Young Adults With Asthma: Multi-Methods Feasibility Study

Jane Murphy¹, MSc; Jenny McSharry¹, PhD; Lisa Hynes², PhD; Gerard J Molloy¹, PhD

¹School of Psychology, National University of Ireland, Galway, Galway, Ireland
²Croí, The West of Ireland Cardiac Foundation, Galway, Ireland

Abstract

Background: Young adults with asthma often report low adherence to inhaled corticosteroids (ICS), leading to uncontrolled symptoms and poor disease outcomes. Technology-enabled digital supports such as mobile health (mHealth) asthma smartphone apps have the potential to support adherence to ICS and asthma self-management. There is a need for feasibility studies to determine the usability, acceptability, and feasibility of these interventions. In addition, it is essential to determine the feasibility of recruiting and retaining young adults to plan future efficacy and effectiveness trials and therefore, establish evidence-based asthma apps.

Objective: This study aimed to determine (1) the feasibility of recruiting and retaining young adults to a trial and (2) the usability, acceptability, and feasibility of using the AsthmaMD app to support adherence to ICS in a population of young adults living with asthma.

Methods: A multi-methods feasibility study was conducted. Young adults aged 18-30 years with asthma and current prescription for ICS were eligible and invited to take part through a university circular email, social media, and general practice sites. Participation involved completing a baseline self-report questionnaire, downloading and using the AsthmaMD app for 2 weeks, and completing the follow-up assessment, including self-report and open-ended questions about participants’ experience of using the app. Primary outcomes included participant recruitment and retention and the usability, acceptability, and feasibility of using AsthmaMD. Quantitative self-report data were analyzed using descriptive statistics, and qualitative open-ended data were analyzed using inductive reflexive thematic analysis.

Results: A total of 122 young adults (females, n=101, 82.8%) with a mean age of 24.4 (SD 3.8) years were recruited and they completed baseline measures. Of the 122 young adults, 59 (48.4%) completed the study. The AsthmaMD app received a mean score of 63.1/100 (SD 20.1) on the System Usability Scale (ie, a standardized measure of usability for technology-based apps), and an overall user satisfaction score of 5.8/10 (SD 2.2). Of the 59 participants who completed the study, 49 (83%) participants used the app ≥1 day per week. Two main themes were identified in the qualitative analysis of user experiences: (1) learning how to use the app to suit the individual and (2) benefits and relevance of using the app.

Conclusions: The findings from this study indicate that it is feasible to recruit and retain young adults to examine efficacy and effectiveness in a future trial and that young adults living with asthma may find AsthmaMD to be usable, acceptable, and feasible to support adherence to ICS. Our findings also identified opportunities to further optimize the usability of AsthmaMD and similar apps. Based on our findings, we recommend providing more accessible information on how to use the app and replacing medical terminology with simplified language within the app to improve usability.

Trial Registration: ISRCTN Registry ISRCTN11295269; https://www.isrctn.com/ISRCTN11295269

(JMIR Form Res 2021;5(9):e28784) doi:10.2196/28784
KEYWORDS

asthma; young adult; medication adherence; self-management; mobile applications; mHealth; intervention; usability; acceptability; feasibility; multi-methods; mobile phone

Introduction

Background

Asthma is a major global health concern with increasing prevalence currently affecting up to 339 million people [1]. More than 60% of adults living with asthma have uncontrolled symptoms [2,3], leading to lower quality of life and higher productivity loss, health care utilization, and asthma mortality [4-7]. Asthma can be effectively controlled through patient engagement in self-management behaviors such as symptom monitoring, avoiding triggers, and adherence to appropriate treatments [8]. According to the Global Initiative for Asthma guidelines [9], adherence to inhaled corticosteroids (ICS), which are considered the most effective long-term asthma control medication [10-13], is essential for the effective management of asthma. Adherence to ICS has been consistently associated with asthma control, improved lung function, reduced symptom exacerbations, and thereby reduced burden of asthma on health care systems [14-17]. However, low adherence remains a major barrier to optimal asthma control, most notably in younger populations [18,19].

A recent review estimated that only 28% of young adults may be fully adherent to their ICS [20]. Adherence to ICS in young adults is further complicated as they become responsible for their own asthma self-management, with parents or caregivers potentially having less involvement in these tasks [21]. This alone constitutes a vulnerable transitional period for young adults [22]. Additionally, during this period, young adults often enter a psychological development phase, coined as emerging adulthood [23,24]. Emerging adulthood has been proposed as a distinct stage of development between adolescence and adulthood from approximately the age of 18 years to 29 years, where individuals typically experience greater autonomy, explore opportunities in education, work, residence, and relationships [23,24], and engage in more health risk behaviors [25-28]. Asthma self-management is affected by this unstable development period. Furthermore, engaging young adults in self-management research often proves challenging in asthma [29] and other chronic conditions [30,31]. Therefore, this population requires attention and there is a clear need to establish appropriate supports to improve self-management, especially adherence to ICS, at this challenging point in the lifespan.

Smartphones provide an existing intervention platform for young adults, given their near universal ownership and high daily use in this population [32]. Although growth in their ownership has not always been equal within or between countries, these devices are now widely used across socioeconomic groups, with young adults being the most likely cohort to own a smartphone in both high and low- and middle-income countries [32]. Moreover, there has been an exponential growth of commercially available mobile health (mHealth) apps, including asthma self-management apps, some of which have demonstrated potential to improve adherence and disease control [33,34]. A recent systematic review [35] identified AsthmaMD as one of the currently available asthma apps with the highest number of evidence-based behavior change techniques (BCTs; n=10) [36] and highest mobile app rating score (score 4.23/5). Additionally, recent qualitative work has explored the technology preferences of young adults with asthma to support adherence to ICS (J Murphy, MSc, under review, November 2020). Young adults’ preferred type of technology to support adherence was a smartphone app, and accordingly, an app is the focus of this study. Based on their preferences for individual app features, we selected AsthmaMD as a suitable, freely available asthma app for this population. However, it is not yet known whether young adults consider this app easy-to-use, acceptable, and feasible, which are essential to ensure successful uptake and use [37,38].

Research Questions

The aim of this study was to assess the feasibility of recruiting and retaining young adults to a trial and the usability, acceptability, and feasibility of using the AsthmaMD app to support adherence to ICS in a population of young adults living with asthma in Ireland. The research questions in this study are as follows:

1. Can participants be recruited to take part in the study? (feasibility of young adult recruitment)
2. Can participants be retained in the study until completion? (feasibility of young adult retention)
3. Do participants find AsthmaMD easy to use? (usability of AsthmaMD)
4. Are participants satisfied with using AsthmaMD to support adherence to ICS and with their overall user experience with the app? (acceptability of AsthmaMD use)
5. Do participants use and would they continue to use AsthmaMD to support adherence to ICS? (feasibility of AsthmaMD use)

Following guidance on conducting feasibility studies [39], the research team established Go/No Go progression criteria to determine if the findings of this study indicated that it would be feasible to examine the efficacy and effectiveness of AsthmaMD in a future randomized controlled trial.

Methods

Design

A multi-methods feasibility design [40] with a 2-week follow-up was employed for this study. The intervention duration was based on input from public and patient involvement (PPI) contributors who felt this was an appropriate duration to obtain a sufficient user experience and that their engagement with the app over this period would accurately indicate their long-term use of the app. This 2-week follow-up is consistent with similar feasibility studies of eHealth and mHealth interventions such as apps for medication adherence [41] and self-management...
This study was reported in accordance with all relevant principles of the extended CONSORT (Consolidated Standards of Reporting Trials) checklist of information to include when reporting a pilot/feasibility trial [47], the CONSORT-EHEALTH checklist, for improving and standardizing evaluation reports of web-based and mHealth interventions [48], the Template for Intervention Description and Replication (TIDieR) checklist [49] to ensure completeness of reporting and replicability of the intervention, and the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [50]. The completed checklists are shown in Multimedia Appendix 3, Multimedia Appendix 4, Multimedia Appendix 5, and Multimedia Appendix 6, respectively.

**Go/No Go Progression Criteria**

In line with relevant guidance [40], the research team identified the key uncertainties that needed to be assessed in relation to the feasibility of using AsthmaMD and examining its efficacy in a future trial in young adults. These key uncertainties were identified from the relevant literature and research team expertise in relevant parameters of feasibility studies. Following discussion, the research team agreed the Go/No Go progression criteria for the study and how these criteria would be interpreted accordingly. Criteria and thresholds were based on recommended sample size for feasibility studies [51], attrition rates in similar studies in young adults [41,42,52,53], validated usability scales and cut-off scores [54,55], and previously used measures and definitions of acceptability [56,57] and feasibility of app use [56-60]. The rationale for the selected thresholds is further outlined in Multimedia Appendix 7. The progression criteria and thresholds for each are presented in Table S1 of Multimedia Appendix 8. The scoring key was devised based on a recent precedent [58]. The scoring key was as follows:

1. If all criteria are rated as green, the trial to test the effectiveness of the app can immediately proceed.
2. If some criteria are partially met and rated as amber, the trial of the app can proceed once relevant criteria have been reviewed and a plan to make relevant amendments has been agreed by the research team.
3. If one criterion is not met and rated as red, the trial can only proceed if (1) no more than 2 criteria are rated red and (2) a plan to make the required amendments has been agreed by the research team.

**PPI**

A total of 4 (2 females) young adult PPI contributors in the age range of 19-27 years, living with asthma, and prescribed ICS were recruited through social media and advertisements in the university campus. They were invited to attend 1 meeting lasting approximately for 1.5 hours. Each contributor was offered a €38 (US $1=€0.85) retail voucher for their contributions. This value was based on INVOLVE guidelines [61], the United Kingdom’s national advisory group to support public involvement in the National Health Service research. The researcher held 2 meetings, each with 2 young adult contributors. The researcher presented the proposed research and contributors were asked to review all study documents individually, including the study information sheet, consent form, invitation letter from a general practitioner (GP), baseline and follow-up measures, and then to discuss their reviews with the group. Specifically, the researcher asked the contributors to review all the documents for clarity and relevance from a lived experience perspective. The appropriate study follow-up duration and recruitment strategies were then discussed among the group. According to a recent PPI Involvement Matrix [62], the contributors fulfilled the cothinks and advisor role at the implementation stage of this research. Appropriate and feasible amendments were made to the study design and documents following each meeting.

Examples of PPI input into the study design and amendments include (1) the 2-week follow-up duration based on the reasons outlined in the design section, (2) using a different font such as Times New Roman in all study documents, as this is often the required font for college-related submissions and can remind young adults of academic work, which may prevent engagement with the study information, (3) providing a description and common examples of oral corticosteroids to increase clarity when asking participants if they have been treated with these in the past year, and (4) increasing the use of terms such as “novel,” “innovative,” and “unique” when describing AsthmaMD as a potential method to support adherence, stimulate young adult interest, and emphasize the app as an alternative to traditional clinical interventions.

**Participants and Recruitment**

Eligible young adults were aged 18-30 years, with a self-reported asthma diagnosis and currently prescribed a form of ICS. Young adults were defined as adults aged 18-30 years to include the emerging adulthood age range of 18-29 years [23,24] and the range of definitions across the health care transition literature [63,64]. This definition has also been used previously in a similar context [65]. Participants were not compensated or offered any incentive to take part in the study. They were invited to take part through an email sent from a weekly university circular to all registered students. The study was also advertised via social media from several accounts including the study’s own Facebook, Twitter, and Instagram accounts, the Asthma Society of Ireland’s Twitter and Facebook, and a range of other community and health-based social media accounts. Additionally, the 4 PPI study contributors shared the study advertisement via their social media. Four GPs in the Galway region also supported study recruitment. The GPs identified all eligible patients on their registers, following a
search of 18-30-year-old adults who were coded with an asthma diagnosis and prescription for ICS. Collectively, they identified 85 eligible patients who were posted a study invitation. As an incentive to support recruitment, GPs were offered an Irish Medical Council–eligible audit template of clinically relevant Global Initiative for Asthma guidelines [9] that was conducive to fulfilling their annual audit requirements.

Sample Size
A minimum of 59 participants were required to complete this study. This target was based on a commonly applied recommendation from the literature, which states that if a problem with 5% probability exists in a potential study participant, then it should be identified in a sample of 59 participants [51]. Based on attrition rates in similar feasibility studies with a 2-week follow-up [41,42,52] and a recent pilot study of an asthma self-management app in a population of young people [53], we anticipated a 25% rate of attrition in this study. To allow for this, we aimed to recruit a minimum of 74 participants for this study.

Outcome Measures
Primary outcomes were the feasibility of participant recruitment and retention, and the usability, acceptability, and feasibility of AsthmaMD. Additional measures were used to collect information on demographics, asthma characteristics, adherence to ICS, asthma control, and smartphone and app use. All baseline and follow-up measures are included in Multimedia Appendix 9 and Multimedia Appendix 10, respectively. Baseline demographic and asthma measures included age, gender, education, ethnicity, eligibility for free or reduced-rate medical treatment and GP services, number of years since asthma diagnosis, emergency department visits, hospital admissions, and treatment with oral steroids in the past year. Adherence to ICS was measured at baseline and at follow-up by using the Medication Adherence Report Scale for Asthma (MARS-A) [66], a 10-item self-report measure. Responses ranged from “(1) Always” to “(5) Never,” where higher scores indicate better adherence. Responses were summed to yield a total score from 10 to 50. The MARS-A demonstrated reliability and construct and convergent validity [67,68]. Asthma control was measured at baseline and at follow-up by using the Asthma Control Test (ACT) [69]. The ACT is a 5-item questionnaire, to which responses are rated on a 5-point scale and summed to provide a total score ranging from 5 to 25, with higher scores representing better asthma control. An ACT score ≤19 indicates uncontrolled asthma [69]. The ACT is the most validated composite measure of asthma control [70-72]. Additionally, information on smartphone and app use were collected at baseline, including number of years using a smartphone, hours per day using apps, ever used an asthma app, and current use of and awareness of an asthma app.

Feasibility of Participant Recruitment
Feasibility of participant recruitment was determined by the number of participants recruited, that is, the number of participants who provided consent and completed baseline measures. The source of recruitment was also recorded by asking participants where they heard about the study.

Feasibility of Participant Retention
Feasibility of participant retention was determined by the number of participants who completed the follow-up assessment, that is, those who completed the study.

Usability of AsthmaMD
The perceived usability of AsthmaMD was determined using the System Usability Scale (SUS) [54] at follow-up. The SUS consists of 10 items to which responses are made on a 5-point scale ranging from “(1) strongly disagree” to “(5) strongly agree”. Scoring of the SUS involves recoding responses. Responses to the odd-numbered items (1, 3, 5, 7, and 9) are subtracted by 1, and responses to the even-numbered items (2, 4, 6, 8, and 10) are subtracted from 5. These recoded responses are then summed, and the sum is multiplied by 2.5 to generate a composite SUS score ranging from 0 to 100, with higher scores indicating greater usability. A score >68 is considered above average [55]. The SUS has demonstrated reliability and convergent and discriminant validity across a range of populations [55,73,74]. It is technology agnostic [75] and therefore remains applicable to a range of technologies as they continuously evolve, such as smartphone apps. Participants were also asked if the app was easy to use, which individual app features they used, and which features they found useful.

Acceptability of AsthmaMD Use
Acceptability of AsthmaMD use was determined at follow-up from participants’ ratings of their overall user satisfaction on a scale of 1-10, whether the app increased adherence awareness and adherence behavior, increased confidence, reduced stress in managing ICS, and if notifications were annoying. Participants’ willingness to recommend the app to another person with asthma and to pay for the app were also measured. These measures are modified from previous feasibility studies of mHealth interventions [56,57], one of which defined acceptability of app use as ≥30% of participants agreeing to similar questions [57].

Feasibility of AsthmaMD Use
Feasibility of app use was determined at follow-up from the self-reported number of days per week participants used the app and their intention to continue to use the app beyond their participation in the study. Participants’ duration of app use was also measured by the self-reported number of minutes per day they used the app. These are frequently used questions in feasibility studies of mHealth interventions [56-60].

Open-ended Questions
Participants were asked 5 open-ended questions at follow-up, which aimed to elicit additional feedback and elaboration of their experience with using the app. These questions were primarily adapted from a recent similar study [53] and included what participants liked the most and the least about the app, any difficulties experienced, suggestions for improvements, and any other comments.

Intervention
AsthmaMD was created by Dr Sam Pejham, a physician and researcher from the University of California, San Francisco.
(USSF) Medical School Clinical Faculty. The app was awarded the UCSF Collaborative Research Network grant, which is a primary care practice-based research network grant. A recent review identified the following BCTs in AsthmaMD: (1) information about health consequences, (2) salience of consequences, (3) information about others’ approval, (4) instruction on how to perform a behavior, (5) demonstration of the behavior, (6) self-monitoring of behavior, (7) self-monitoring of outcome(s) of behavior, (8) feedback on behavior, (9) feedback on outcome(s) of behavior, and (10) prompt/cues. The first author (JM) also screened and coded AsthmaMD for the presence of BCTs by using the BCT taxonomy (BCTTv1) [36]. Consistent with Ramsey et al [35], JM identified the presence of these 10 BCTs, with the addition of credible source in each video tutorial as Dr Pejham delivers the information and either speaks in favor of the relevant self-management behavior or using AsthmaMD to support these behaviors. The app was available to download for free from Google Play Store (Android) and App Store (iPhone). Screenshot of the AsthmaMD user interface and its individual features can be seen in Multimedia Appendix 11.

AsthmaMD allows users to log their peak expiratory flow meter readings, which can indicate response to treatment and triggers for worsening symptoms or determine a baseline for action plans [9]. In addition to peak expiratory flow readings, the app allows users to log their symptoms, triggers, medications, and other notes in the form of a diary and share these data with their physicians or other persons. It provides a line chart of peak expiratory flow meter readings and symptoms, indicating the severity of each, allows users to create customized reminders to take their medication, and guides users through their asthma action plan. Finally, the app provides video tutorials on using your peak flow meter, understanding asthma and asthma medication, and using AsthmaMD.

Procedure
The procedure involved a baseline and follow-up web-based questionnaire administered by LimeSurvey (version 3.19) [46]. Participants were asked to download AsthmaMD at the end of the baseline questionnaire. They were not instructed on how, when, how often, or how long to use AsthmaMD. It was decided to leave the frequency and duration of use to the users’ discretion as people mainly use these technologies independently and autonomously [38]; thus, this may provide a more accurate indication of how participants would use the app in their daily life after study participation, that is, their long-term use of the app. Participants could tailor the app to their own personal regime at their discretion, for example, by entering the name and dosage of their medications at any point during the 2 weeks. No other physical or informational materials were provided to participants as part of the intervention. Participants were encouraged to make themselves familiar with the app’s data privacy policy before downloading the app. If participants were interested in completing the follow-up questionnaire, they were asked to provide their email address and they received an email with the link to the follow-up study 2 weeks later from the researcher, a PhD candidate in health psychology. If these participants had not completed the follow-up questionnaire within 3 working days, they were sent a total of 3 successive reminder emails every 3 working days.

Data Analysis
Quantitative analysis was performed using SPSS software (version 26, IBM Corp) [76]. Descriptive statistics were used to analyze baseline data and primary study outcomes. Independent two-tailed t tests and χ2 tests of association were used to examine the differences between participants who completed the study and those who did not, that is, those lost to follow-up. Cronbach α was conducted to examine the internal consistency of the multi-item scales employed.

Qualitative analysis was performed using NVivo software (version 12, QSR International) [77]. An inductive, reflexive thematic analysis [78,79] was conducted on participant responses to open-ended questions in the follow-up assessment. The analysis followed the 6 stages of Braun and Clarke’s [78] proposed guidance: (1) familiarizing oneself with the data, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) producing the report with illustrative quotes. The first author engaged in data immersion by repeatedly reading and coding all data. Initial codes were collated into potential themes relevant to the complete data set. Analysis was iterative with codes, themes, and subthemes refined throughout.

Results
Baseline Measures
Table 1 presents the information collected at baseline. The mean age of the sample was 24.4 (SD 3.8) years, and the majority were females of White Irish ethnicity who had at least tertiary or higher-level education and were living with asthma for at least 10 years. On average, the sample reported low adherence to ICS and uncontrolled asthma at baseline and follow-up assessments. The majority were using a smartphone for 5-10 years, they spent at least 3 hours per day using smartphone apps, had not previously and were not currently using an asthma app, and were unaware of asthma apps. Table 1 also provides a comparison of the baseline measures between participants who completed the study and those who did not. Participants who completed the study were significantly older, and a significantly greater proportion had postgraduate education and were aware of an asthma app in comparison to those who did not complete the study. No other significant differences in baseline measures existed between those who did and did not complete the study. Cronbach α for the MARS-A at baseline (α=.83) and follow-up (α=.83), for the ACT at baseline (α=.82) and follow-up (α=.83) and for the SUS at follow-up (α=.92) were all >.70 [80], demonstrating valid internal consistency for all scales in this study.
Table 1. Baseline measures according to study completion (N=122).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Participants who completed baseline (N=122)</th>
<th>Participants who completed baseline and follow-up (n=59)</th>
<th>Participants who completed baseline only (n=63)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>24.4 (3.8)</td>
<td>25.1 (3.8)</td>
<td>23.6 (3.6)</td>
<td>.03</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.38</td>
</tr>
<tr>
<td>Female</td>
<td>101 (82.8)</td>
<td>47 (80)</td>
<td>54 (86)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>21 (17.2)</td>
<td>12 (20)</td>
<td>9 (14)</td>
<td></td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Second level</td>
<td>18 (14.8)</td>
<td>4 (7)</td>
<td>14 (22)</td>
<td></td>
</tr>
<tr>
<td>Tertiary/higher level</td>
<td>63 (51.6)</td>
<td>31 (52)</td>
<td>32 (51)</td>
<td></td>
</tr>
<tr>
<td>Postgraduate</td>
<td>41 (33.6)</td>
<td>24 (41)</td>
<td>17 (27)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.49</td>
</tr>
<tr>
<td>White Irish</td>
<td>99 (81.1)</td>
<td>51 (86)</td>
<td>48 (76)</td>
<td></td>
</tr>
<tr>
<td>Other White background</td>
<td>13 (10.7)</td>
<td>3 (5)</td>
<td>10 (16)</td>
<td></td>
</tr>
<tr>
<td>Black/Black Irish</td>
<td>2 (1.6)</td>
<td>1 (2)</td>
<td>1 (2)</td>
<td></td>
</tr>
<tr>
<td>Asian/Asian Irish</td>
<td>3 (2.5)</td>
<td>2 (3)</td>
<td>1 (2)</td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td>3 (2.5)</td>
<td>1 (2)</td>
<td>2 (3)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>2 (1.6)</td>
<td>1 (2)</td>
<td>1 (2)</td>
<td></td>
</tr>
<tr>
<td>Medical card holder, Yes, n (%)</td>
<td>25 (20.5)</td>
<td>8 (14)</td>
<td>17 (27)</td>
<td>.07</td>
</tr>
<tr>
<td>General practitioner visit card holder, Yes, n (%)</td>
<td>10 (8.2)</td>
<td>2 (3)</td>
<td>8 (13)</td>
<td>.06</td>
</tr>
</tbody>
</table>

| Recruitment source, n (%)     |                                             |                                                          |                                               | .08                 |
| Facebook                      | 29 (23.8)                                   | 15 (25)                                                  | 14 (22)                                       |                     |
| Twitter                       | 23 (18.9)                                   | 10 (17)                                                  | 13 (21)                                       |                     |
| Instagram                     | 22 (18.0)                                   | 9 (15)                                                   | 13 (21)                                       |                     |
| University student mailer     | 23 (18.9)                                   | 7 (12)                                                   | 16 (25)                                       |                     |
| Word of mouth                 | 23 (18.9)                                   | 17 (29)                                                  | 6 (9)                                         |                     |
| General practitioner/nurse    | 2 (1.6)                                     | 1 (2)                                                    | 1 (2)                                         |                     |

| Asthma characteristics        |                                             |                                                          |                                               | .14                 |
| Years since diagnosis, n (%)  |                                             |                                                          |                                               |                     |
| <1 y                          | 1 (0.8)                                     | 0 (0)                                                    | 1 (2)                                         |                     |
| 1-2 y                         | 9 (7.4)                                     | 5 (8)                                                    | 4 (6)                                         |                     |
| 2-5 y                         | 14 (11.5)                                   | 8 (14)                                                   | 6 (10)                                        |                     |
| 5-10 y                        | 15 (12.3)                                   | 8 (14)                                                   | 7 (11)                                        |                     |
| 10+ y                         | 76 (62.3)                                   | 38 (64)                                                  | 38 (60)                                       |                     |
| Don’t know                    | 7 (5.7)                                     | 0 (0)                                                    | 7 (11)                                        |                     |
| Visited emergency department in the last year, Yes, n (%) | 13 (10.7) | 8 (14) | 5 (8) | .32 |
| Hospital admission in last year, Yes, n (%) | 8 (6.6) | 5 (9) | 3 (5) | .41 |
| Prescribed oral steroids in last year, Yes, n (%) | 46 (37.7) | 26 (44) | 20 (32) | .16 |
| Asthma control, n (%)         |                                             |                                                          |                                               | .22                 |
| Uncontrolled asthma: Asthma Control Test score ≤19 | 64 (52.5) | 28 (47) | 36 (57) |                     |
| Controlled asthma: Asthma Control Test score >19 | 58 (47.5) | 31 (53) | 27 (43) |                     |
| Total Asthma Control Test score, mean (SD), range | 19.0 (4.2), 7-25 | 19.2 (4.4), 7-25 | 18.8 (4.0), 9-25 | .59 |
Primary Outcomes

**Participant Recruitment**
A total of 122 participants were recruited to this study and they completed baseline measures from September to December 2020. In relation to where the participants heard about the study, Facebook was the most common source, followed by Twitter, university student mailer, word of mouth, Instagram, and GP or nurse.

**Participant Retention**
Of the 122 participants who completed the baseline measures, 59 (48.4%) completed the follow-up assessment, that is, were retained until study completion. Information regarding participant flow throughout the study is shown in Multimedia Appendix 12. Of the participants who completed the study, word of mouth was the most common recruitment source, followed by Facebook, Twitter, Instagram, university student mailer, and GP or nurse.

**Usability of AsthmaMD**
The mean SUS score for AsthmaMD was 63.1 (SD 20.1). A score >68 indicates average usability. A total of 25 (46%) participants “agreed/strongly agreed” that the app was easy to use. Table 2 presents the proportion of participants who used individual AsthmaMD features and who found these useful. The feature that was most used and most frequently reported as useful was the symptom log. The feature that was least used and least frequently reported as useful was the forced expiratory volume-1 log.
Table 2. AsthmaMD features, behavior change techniques present, and proportion of participants who used and found each feature useful (n=59).a

<table>
<thead>
<tr>
<th>App feature</th>
<th>Behavior change technique</th>
<th>Participants who used feature, n (%)</th>
<th>Participants who found feature useful, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptom log</td>
<td>Self-monitoring of behavior</td>
<td>41 (70)</td>
<td>30 (73)</td>
</tr>
<tr>
<td>Log of medication use</td>
<td>Self-monitoring of behavior</td>
<td>39 (66)</td>
<td>18 (46)</td>
</tr>
<tr>
<td>Trigger log</td>
<td>Self-monitoring of behavior</td>
<td>37 (63)</td>
<td>29 (78)</td>
</tr>
<tr>
<td>Peak flow log</td>
<td>Self-monitoring of behavior outcome(s)</td>
<td>13 (22)</td>
<td>10 (77)</td>
</tr>
<tr>
<td>Oximetry log</td>
<td>Self-monitoring of behavior outcome(s)</td>
<td>3 (5)</td>
<td>2 (67)</td>
</tr>
<tr>
<td>Forced expiratory volume-1 log</td>
<td>Self-monitoring of behavior outcome(s)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Notes</td>
<td>Self-monitoring of behavior</td>
<td>5 (9)</td>
<td>2 (40)</td>
</tr>
<tr>
<td>Diary</td>
<td>Feedback on behavior</td>
<td>26 (44)</td>
<td>17 (65)</td>
</tr>
<tr>
<td>Line chart</td>
<td>Feedback on behavior outcome</td>
<td>26 (44)</td>
<td>17 (65)</td>
</tr>
<tr>
<td>Send report to physician/other</td>
<td>Information about other’s approval</td>
<td>3 (5)</td>
<td>2 (67)</td>
</tr>
<tr>
<td>Reminders</td>
<td>Prompts/cues</td>
<td>23 (39)</td>
<td>18 (78)</td>
</tr>
<tr>
<td>Action plan</td>
<td>Instruction on behavior performance</td>
<td>12 (20)</td>
<td>7 (58)</td>
</tr>
<tr>
<td>Video tutorial on using your peak flow meter</td>
<td>Instruction on behavior performance</td>
<td>3 (5)</td>
<td>2 (67)</td>
</tr>
<tr>
<td>Video tutorial on understanding asthma</td>
<td>Information about health consequences</td>
<td>9 (15)</td>
<td>7 (78)</td>
</tr>
<tr>
<td>Video tutorial on asthma medication</td>
<td>Information about health consequences</td>
<td>6 (10)</td>
<td>5 (83)</td>
</tr>
<tr>
<td>Video tutorial on AsthmaMD</td>
<td>Instruction on behavior performance</td>
<td>5 (9)</td>
<td>3 (60)</td>
</tr>
<tr>
<td>AsthmaMD frequently asked questions</td>
<td>Instruction on behavior performance</td>
<td>6 (10)</td>
<td>1 (17)</td>
</tr>
</tbody>
</table>

aThe number of participants who used each feature and who found each feature useful varied per individual feature.

Acceptability of AsthmaMD Use

AsthmaMD received a mean score of 5.8 (SD 2.2) for participants’ overall experience in using the app for managing their adherence to ICS. Of the 59 participants, 34 (58%) participants rated their experience ≥5/10. In total, 27 (46%) participants “agreed/strongly agreed” that the app increased their awareness of their adherence, and 21 (36%) “agreed/strongly agreed” that the app increased their adherence to ICS. In response to whether the app increased their confidence in managing their ICS, 18 (31%) participants “agreed/strongly agreed.” A total of 15 (25%) participants “agreed/strongly agreed” that the app reduced the stress in managing their adherence to ICS, and 25 (42%) “agreed/strongly agreed” that the app notifications did not annoy them. A total of 25 (42%) participants reported that they would recommend the app to another person with asthma and 14 (24%) were “not sure.”

Feasibility of AsthmaMD Use

In total, 49 (83%) participants reported using AsthmaMD ≥1 day per week. The mean number of days per week participants reported using the app was 3.2 (SD 1.9). In response to whether participants would continue to use the app after their participation in the study, 16 (27%) said “yes” and 14 (24%) were “not sure.” In relation to number of reported minutes per day participants used the app, 26 (44%) used it for less than 5 minutes, 22 (37%) used it for 5-10 minutes, 7 (12%) used it for 10-15 minutes, and 4 (7%) used it for 15-20 minutes per day.

Qualitative Findings From Open-ended Questions

Of the 59 participants, 40 (68%) participants responded to the open-ended questions. Two main themes were identified in the data: (1) learning how to use the app to suit the individual and (2) benefits and relevance of using the app.
Learning How to Use the App to Suit the Individual

Participants had a range of experiences with learning how to use AsthmaMD and making it personally useful. Some discussed how easy the app was to use on first use, for individual features and overall.

...It was easy to input my personal data and then to track symptoms and triggers. [Participant 47, male, age 25 years, tertiary/higher level education]

Other participants felt the app required an investment of time for "trial and error" (Participant 23, female, age 30 years, tertiary/higher level education) in the beginning to learn how to use it and subsequently to understand how it could be useful. This initial investment was worthwhile for some who progressed in using the app thereafter and found individual features useful that they may otherwise not have used.

...It took me a while to fully understand the proper function of the app...to realize what the action plan was and what it was for...Once I understood it made more sense to me why I would use an app like this...I will be looking into getting an action plan in place with my GP which I wouldn’t have thought of before downloading the app. [Participant 39, male, age 27 years, postgraduate education]

However, despite this time investment, some participants continued to find the app difficult to use. Two factors contributed to this experience. First, participants commented on the unnoticeable location of features such as the AsthmaMD app video tutorial, which may have enabled participants to more efficiently and easily learn how to use the app if it had been more apparent to the user.

...I couldn’t see where to access videos/tutorials to help with using the app which would’ve been useful. [Participant 14, female, age 20 years, second-level education]

Participants suggested solutions to this, including relocating the information and features on how to use AsthmaMD that already exist within the app to a more obvious location that is easier for users to find and adding more information on how to use individual features, immediately presenting this to users on their first use of the app. Specifically, suggestions included adding a help section within each feature on how and why to use it.

...Help sections within each section: eg, when I click on "action plan" it would be useful if there was a link to help with that - Other apps have a question mark at the top which when you click it tells you about the section, eg, directions on how to set it up and when and why it would be helpful...I would also include information about what the action plan is and how to go about getting in with the GP” [Participant 11, female, age 29 years, postgraduate education]

Second, participants felt there was too much medical jargon used in the app. They felt the app assumed user knowledge of terminology, which they did not have, thereby making it difficult to understand.

...I wasn’t familiar with some of the terms...the app just presumes you know things. [Participant 57, female, age 21 years, tertiary/higher level education]

Benefits and Relevance of Using the App

This theme includes 2 subthemes: (1) prompt to action and (2) useful for me or for others.

Prompt to Action

Participants described how their use of the app resulted in increased awareness of their asthma self-management, which often resulted in increased self-management behavior, including improved adherence to ICS and seeking a GP consultation to track peak flow and devise an asthma action plan.

...this app has highlighted areas of my asthma that I wasn’t aware of. I have hooked in to my GP to get some answers and to be able to track the peak flow. [Participant 41, female, age 26 years, tertiary/higher level education]

One participant discussed how the app identified their asthma as suboptimally controlled, which prior to their use of this app they perceived as well-controlled. This objective indication of asthma control may have otherwise remained unknown to the participant who now intends to improve their self-management as a result.

...Reminded me about some of the measures to control your asthma. I thought I managed mine well, but the app told me it wasn’t so I have to improve on it and I definitely will, due to the app! [Participant 62, male, age 30 years, tertiary/higher level education]

In particular, the reminder to take ICS was discussed as a useful feature to overcome forgetfulness, which was noted as a barrier to taking ICS, and thereby improve adherence to this medication. However, 1 participant felt the regular reminders unnecessarily increased their awareness of taking their ICS and made the behavior feel more effortful.

...It made taking my inhaler into a chore. I’m aware that it is a chore but the constant notifications annoyed me. [Participant 54, male, age 19 years, second-level education]

Useful for Me or for Others

Participants discussed the individual features they found and would find especially useful based on their personal self-management needs and preferences at different times.

...This would be very handy when I need to track my peak flow. [Participant 34, female, age 28 years, tertiary/higher level education]

...I used this app to track my peak flow, I take my inhaler properly every day so did not need it to help me with this. [Participant 50, female, age 28 years, postgraduate education]

Others expressed how they did not find the app particularly useful owing to a lack of perceived asthma severity, already having a daily routine for taking ICS or already using a reminder on their phone, and perceiving no need to use additional technology features.

https://formative.jmir.org/2021/9/e28784

JMIR Form Res 2021 | vol. 5 | iss. 9 | e28784 | p.11
(page number not for citation purposes)
...the app would be good for someone that isn’t compliant with their medications however I am so used to taking it now I don’t need it. [Participant 12, female, age 26 years, tertiary/higher level education]

A common perception among these participants was that although the app may not be useful for them on a personal level, they felt it would be useful for other young adults living with asthma such as those with low adherence to ICS, more severe or frequent symptoms, or those newly diagnosed with the condition.

...this might be more useful for moderate/severe asthma patients who are struggling to determine their disease pattern. I can see the benefit of the app for this type of patient. It seems less useful for controlled asthma, except for the reminders. [Participant 21, female, age 30 years, postgraduate education]

Participants proposed suggestions to increase the personal relevance of the app by adding all ICS treatments to the medication list within the app, adding options to set reminders for additional self-management behaviors such as logging symptoms and visually presenting the relationship between users’ data on symptoms and medication use in a chart.

...a flow chart of your symptoms and it’s correlation with medication use...that would be useful. [Participant 2, female, age 24 years, postgraduate education]

Discussion

Principal Findings

The findings from this study indicate that the AsthmaMD app is usable, acceptable, and feasible as defined by the study progression criteria, to support adherence to ICS in a population of young adults living with asthma, and that it appears feasible to recruit and retain young adults to examine its effectiveness in a prospective randomized controlled trial. Table S2 of Multimedia Appendix 13 presents the outcome of each Go/No Go progression criterion. All criteria were fully met (Green: Proceed), except for the usability of AsthmaMD, which was partially met (Amber: Amend). Based on our findings, we propose 3 strategies to improve the app’s usability.

First, we propose relocating the AsthmaMD app video tutorial to the main app home screen and sending a notification that directs users to this tutorial when they first download and open the app. Second, we suggest adding a help section within each feature that explains how to use it and the benefits of use. Third, we propose removing all medical terminology and simplifying the language used. A recent theory-based reframing approach to address asthma perceptions and ICS treatment beliefs offers an alternative from the traditional medical explanation [81]. This approach describes asthma as a condition in which the lungs are “out of balance” and “overreact” in response to triggers. ICS are described as “natural helpers” that “top up” natural steroids in the body to prevent overreactions and restore balance in the lungs. Initial acceptability has been demonstrated in young adults who perceived the approach as coherent and easy to understand [81]. In particular, they appreciated its use of nonmedical jargon and felt the language presented the information in a more positive manner [81]. Perhaps AsthmaMD could employ the language used in this reframing approach to increase its usability.

The progression criteria for the feasibility of participant recruitment and retention were fully met; however, the apparent rate of attrition (63/122, 51.6%) was twice the anticipated rate (31/122, 25.4%). However, it must be noted that only 63.9% (78/122) of the baseline sample provided their email address to receive and complete the follow-up assessment. Therefore, it may be argued that it was only possible for this proportion of participants to complete the study, meaning the study had an attrition rate of 24% (19/78). After completing the baseline questionnaire, the following page thanked participants for taking part and asked those who wished to complete the follow-up assessment to follow a link and provide their email address. This is a necessary ethical measure to maintain confidentiality by ensuring email addresses are not linked to any other data provided by participants. Although participants may have chosen not to provide their email address at this point, it is likely that some did not finish reading this page and therefore unintentionally did not provide their email address to take part in the follow-up. Asking participants to provide their email address at an earlier point in the baseline questionnaire could help overcome this issue in future research.

Nonetheless, we recommend that future trials being conducted with a similar design factor in an attrition rate of at least 24% over 2 weeks. It is important to recognize that attrition rates are likely to be lower where there is direct participant contact as this may enhance engagement and commitment to the study through the establishment of a researcher-participant relationship [82]. The importance of this personal connection is reflected in the findings with word of mouth being the most effective way of retaining participants. The research team may increase the use of word of mouth by attending and providing study information at university lectures, sports club training events, fundraising events, and Asthma in the Pharmacy Days (events at which an asthma nurse offers patient education on asthma in designated pharmacies nationwide). Word of mouth was followed closely by Facebook as an effective strategy of recruiting and retaining participants. Therefore, we also recommend that future trials advertise the study on all existing Facebook accounts in addition to other social media platforms of the organizations supporting recruitment. Regardless of the organization’s most active platform, it appears that study advertisements on Facebook engages young adults most likely to complete the study.

To determine the feasibility of AsthmaMD use, the frequency of app use during the study period was considered. This was measured by the self-reported number of days per week that participants used the app. The research team selected using the app 1 day per week or more as the “Green: Proceed” threshold for this criterion. This relatively infrequent use in certain contexts may raise questions about the potential potency of this as a behavior change support for self-management of this chronic condition. However, asthma has a symptomatic/asymptomatic nature and varies over time often due to patient adherence, physical activity, allergen or irritant...
exposure, seasonal changes, or viral respiratory infections [9]. Therefore, frequency of app use will likely vary depending on the users’ needs and preferences at different times. It may be relevant to gauge the use and usefulness of AsthmaMD and similar apps in line with these patient needs such as reducing asthma symptoms for example. This is consistent with recent findings indicating that patients with asthma may keep a self-management app on their phone despite infrequently using it, in order to potentially use it when needed, to manage increasing symptoms [83].

This is line in with the concept of effective engagement with digital behavior change interventions, defined as adequate engagement with the intervention to achieve intended outcomes [84]. This concept proposes that intervention use alone does not determine health behavior outcomes as these may be influenced by additional factors such as motivation and self-regulation skills. Additionally, users may value alternative outcomes to those intended by the developers. Effective engagement is defined based on the aim of the specific intervention and must be identified within the context of that intervention. Evidence of effective engagement has recently been found in the context of an app for medication adherence in adolescents with asthma [85]. Overall app use was not associated with a difference in adherence; however, use of specific features such as the health care provider chat significantly increased adherence behavior. Therefore, if the efficacy of AsthmaMD is established in future trials, it may be more valuable to identify and encourage effective engagement as opposed to simply more engagement with the app.

The jobs-to-be-done theory [86] provides a similar potential theoretical explanation of this. This framework of consumer action proposes that people use a product or service because it meets a need to get a job done, not for the sake of the product or service itself [87]. Patients may want to reduce symptoms when they occur and may use an asthma app for this purpose. Success is determined by getting the job done [87]. Therefore, when symptoms are eliminated, patient users may feel the app has served its purpose and not perceive a need to continue app use. Although health care providers and behavioral scientists may regard constant patient engagement with self-management supports as ideal, this may not be reasonable or necessary. This threshold of using AsthmaMD at least 1 day per week was exceeded as participants used the app, on average, 3 days per week. In total, 51% (30/59) of the participants reported that they would or were unsure if they would continue to use the app. Improving its usability through the above suggestions may increase intent to continue using AsthmaMD beyond the study.

Additionally, it must be noted that the qualitative findings indicate a common perception that participants feel the app may not be personally useful but would be for others with asthma. This may be due to a lack of perceived asthma severity and ICS necessity beliefs, which have consistently predicted self-management behavior and, specifically, adherence to ICS [88-91]. Furthermore, this perceived personal relevance has been found to influence uptake and engagement with a range of eHealth and mHealth interventions [92] and asthma apps [93]. Therefore, asthma and ICS treatment beliefs may need to be addressed to increase the personal relevance and therefore, use of these interventions to improve suboptimal adherence and self-management in these participants. Horne’s reframing approach [81] aims to modify these perceptions. Given its initial acceptability employing this approach within AsthmaMD may also generate more medically accurate asthma perceptions and ICS beliefs, for example, perceiving asthma as a long-term condition requiring regular ICS medication, and in turn, increasing the personal relevance and therefore, use of the app.

The qualitative findings also indicate a lack of knowledge about certain asthma self-management behaviors such as what is an action plan and the purpose of having this in place. It appears this inadequate knowledge contributed to an initial lack of understanding about how AsthmaMD and specific features could be useful. Perceiving little usefulness or benefits of using the app may prevent user engagement. However, an interest in acquiring this self-management information and engaging in the behavior was apparent. This highlights the need and potential for GPs to assess the knowledge of young adults with asthma and address their needs in consultations such as providing information on asthma action plans and devising the patients’ personal action plans with them. Within these consultations, there is potential for GP advocacy of asthma apps such as AsthmaMD to encourage patient use of these technologies to support self-management. Health care providers have reported significant support for asthma mHealth interventions [94,95] and their professional advocacy of these interventions plays an influential role in patient uptake and use of asthma apps specifically [83] and similar interventions in populations with other chronic conditions [96].

Strengths and Limitations

This study has several strengths and potential limitations. The sample population had a female majority of White Irish ethnicity and with at least tertiary or higher-level education. Of the participants who completed the study, a significantly higher proportion had postgraduate education and were aware of an asthma app and were significantly older than those who did not. However, although significant, this mean age difference was only 1.5 years, which is unlikely, from a psychological perspective, to be developmentally significant at this point in the lifespan. The female majority partly reflects the gender disparities that exist in asthma. In adulthood, females have increased asthma prevalence (9.8% vs 5.5%) [97], hospitalizations, and mortality [98,99]. Furthermore, females are more likely to be educated about asthma control and management than their male counterparts [100,101]. Therefore, they have a higher asthma burden and appear to take increased action concerning their symptoms, which may partially account for higher female engagement in this study. However, adherence and self-management in all adults with asthma requires attention. Both males and females have reported suboptimal asthma control and adherence to ICS [15,102-105]. Therefore, obtaining user feedback with adequate gender balance is important for developing and assessing suitable asthma apps. Further breakdown of recruitment sources by gender identified word of mouth and Twitter as the highest yielding sources of male participants who completed this study. Future studies should consider using these recruitment strategies to target males specifically, for example, providing study information at male
sports club events and advertising via their Twitter accounts. In addition, asthma disproportionately affects racial/ethnic minorities [106] and low socioeconomic groups [107-109]. Therefore, it is essential to engage a diverse sample of young adult users to establish appropriate asthma supports. Snowball recruitment techniques and crowdsourcing may increase diversity in future studies. Additionally, our results are based on a sample of young adults with asthma in a particular geographic and sociocultural context; therefore, further research is needed to support our findings in other contexts.

The high mobility of young adults may have partially accounted for the low recruitment rate from GP letters. In Ireland, 56% of 25-34-year-old adults have received tertiary/higher level education [110]. Young adults often move out of the family home when starting this education and continue to change residence beyond this as they explore career opportunities, cohabitation, or home ownership. Therefore, posting mail to their family home address may not be an effective method of contacting this population. A consequence of this low recruitment from GP was that the sample was primarily recruited through social media and therefore were likely to be more digitally literate and motivated to self-select to participate in self-management research. Future research should explore how best to enroll young adults from GP, including those who are harder to reach.

According to relevant guidance, the design and objectives of feasibility studies should be based on the key uncertainties that exist in relation to the intervention or trial and that a randomized design may not be necessary to address these [40]. Therefore, a nonrandomized design was deemed appropriate for this study. This design is consistent with recent feasibility studies of self-management mHealth interventions for young people with a range of chronic conditions [59,111-115], including asthma [53]. Furthermore, recent randomized feasibility studies of such interventions for young people [116,117] and asthma specifically [44,118] have reported no significant difference in retention rates between intervention and control groups. Therefore, the feasibility of randomizing young adults was not considered a key parameter for this initial investigation in which feasibility of recruitment and retention were deemed more pertinent to focus on at this early stage. However, we recognize that willingness to be randomized is specific to each study context and population and therefore, recommend that future similar studies consider this factor.

Our study was also limited by self-reported app use. We acknowledge that self-report measures often lead to under or overreporting of smartphone use [119-122] and while they can provide useful context [84], they may also lead to biased reporting of engagement in mHealth interventions [123]. However, collecting usage data through the app would have raised substantial ethical and data protection concerns with this study as the research team is not affiliated with the developers of AsthmaMD and so would constitute an additional, external party that would require access to this user data. Therefore, it was not considered worthwhile to pursue this method at this initial stage of app evaluation. However, future studies should consider using these objective measures in combination with self-reported app use that can complement and provide meaningful context to these objective data. We also recommend that future similar studies draw on the Theoretical Framework of Acceptability [124] in designing open-ended questions to capture this concept more extensively.

Finally, the aim of this initial investigation was to determine the usability, acceptability, and feasibility of using the AsthmaMD app as an adherence support. The focus was solely on app use. Accordingly, a 2-week follow-up was deemed appropriate to achieve this aim based on PPI input and supported by similar studies and app retention rates. However, future trials that aim to test the efficacy and effectiveness of AsthmaMD to improve adherence behavior and asthma outcomes should consider a longer follow-up duration that allows time for the clinical effect of ICS to maximize in order to optimally increase asthma control. Although a reduction in inflammation and symptoms can occur within days, it may take ICS several months to reach a plateau and can vary per patient [9-11]. Recent efficacy trials of asthma self-management mHealth interventions have used a 6-month follow-up duration in adults [125,126] and adolescents [127]. This would also provide a representation of longer term app use across this population.

This study also has several strengths, including a near equal proportion of participants with uncontrolled and controlled asthma and a range of adherence to ICS in the baseline and follow-up sample. Therefore, this study obtained user feedback from a sample with symptom control and adherence that is likely representative of the general population of young adults living with asthma. Additionally, PPI was conducted by actively involving the relevant patient group in deciding key aspects of this study design. PPI is valued for several reasons in research. Incorporating the patient lived experience can improve research efficiency by increasing relevance, recruitment, and retention, and benefit meaningful dissemination and implementation [128]. It can also increase the transparency and accountability of research [129]. Therefore, it is likely that PPI enriched the quality and relevance of this study to young adults with asthma. Finally, the Go/No Go progression criteria were developed by a research team with combined expertise on treatment adherence, self-management in young adults, and feasibility studies.

**Conclusion**

The long-term effectiveness of asthma apps such as AsthmaMD in young adults remains unknown. Prior to examining app effectiveness is determining app usability, acceptability, and feasibility in the target user population. AsthmaMD was deemed usable, acceptable, and feasible as defined in this study to support adherence to ICS in a cohort of young adults living with asthma. It also appears feasible to recruit and retain young adults for further research studies, which is critical for conducting prospective trials examining efficacy, effectiveness, and cost-effectiveness. Before proceeding further, we recommend improving the usability of AsthmaMD by providing more of and relocating existing information on how to use the app to a more accessible location and replacing medical terminology with the simplified language used in Horne’s [81] reframing approach to asthma and ICS. Nevertheless, this study has demonstrated potential for successful uptake and use of AsthmaMD and similar smartphone apps. Given the ubiquitous...
use of smartphones, these apps may be a scalable and accessible solution to support adherence to ICS in young adults as they typically experience additional developmental demands throughout this stage of the lifespan. Improving suboptimal adherence to ICS is essential to effective asthma management and to reduce the significant burden of uncontrolled asthma on a personal, social, and economic level.

Acknowledgments
This study is funded by the Irish Research Council, Government of Ireland Postgraduate Scholarship Program (GOIPG/2019/4330). The views expressed are those of the authors and not necessarily those of the Irish Research Council. Permission to use the MARS-A was obtained from Professor Rob Horne, Professor of Behavioral Medicine, School of Pharmacy, University College London. The license to use the ACT was obtained from OptumInsight Life Sciences, Incorporated.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Participant information sheet.
[PDF File (Adobe PDF File), 111 KB - formative_v5i9e28784_app1.pdf]

Multimedia Appendix 2
Study consent form.
[PDF File (Adobe PDF File), 98 KB - formative_v5i9e28784_app2.pdf]

Multimedia Appendix 3
CONSORT (Consolidated Standards of Reporting Trials) checklist for reporting a pilot or feasibility trial.
[PDF File (Adobe PDF File), 90 KB - formative_v5i9e28784_app3.pdf]

Multimedia Appendix 4
CONSORT-eHEALTH (Consolidated Standards of Reporting Trials-eHealth) checklist V1.6.1.
[PDF File (Adobe PDF File), 2361 KB - formative_v5i9e28784_app4.pdf]

Multimedia Appendix 5
The Template for Intervention Description and Replication (TIDieR) checklist for intervention description and replication.
[PDF File (Adobe PDF File), 139 KB - formative_v5i9e28784_app5.pdf]

Multimedia Appendix 6
CHERRIES (Checklist for Reporting Results of Internet E-Surveys) for this study.
[PDF File (Adobe PDF File), 185 KB - formative_v5i9e28784_app6.pdf]

Multimedia Appendix 7
Rationale for Go and No Go progression criteria.
[PDF File (Adobe PDF File), 199 KB - formative_v5i9e28784_app7.pdf]

Multimedia Appendix 8
Go and No Go progression criteria.
[PDF File (Adobe PDF File), 125 KB - formative_v5i9e28784_app8.pdf]

Multimedia Appendix 9
Baseline questionnaire.
[PDF File (Adobe PDF File), 233 KB - formative_v5i9e28784_app9.pdf]

Multimedia Appendix 10
Follow-up questionnaire.
[PDF File (Adobe PDF File), 229 KB - formative_v5i9e28784_app10.pdf]
Multimedia Appendix 11
Screenshots of AsthmaMD user interface.

[PDF File (Adobe PDF File), 413 KB - formative_v59e28784_app11.pdf]

Multimedia Appendix 12
CONSORT (Consolidated Standards of Reporting Trials) 2010 flow diagram.

[PDF File (Adobe PDF File), 118 KB - formative_v59e28784_app12.pdf]

Multimedia Appendix 13
Outcomes of Go and No Go progression criteria.

[PDF File (Adobe PDF File), 150 KB - formative_v59e28784_app13.pdf]

References
   Global%20asthma%20Report%202018%20Embargo.pdf [accessed 2021-01-21]
2. Uncontrolled asthma among adults, 2016. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/asthma/
   asthma_stats/uncontrolled-asthma-adults.htm [accessed 2021-01-21]
3. Easing the economic burden of asthma: The impact of a universal asthma self-management programme, June 2019. Asthma
   Society of Ireland. URL: https://www.asthma.ie/sites/default/files/files/document_bank/2019/Jun/
   ASI%20-%20Easing%20the%20Economic%20Burden%20of%20Asthma%20Report.pdf [accessed 2021-01-21]
   workplace productivity in employed patients who are symptomatic despite background therapy: a multinational survey.
6. Stridsman C, Axelsson M, Warm K, Backman H. Uncontrolled asthma occurs in all GINA treatment steps and is associated
   with worse physical health - a report from the OLIN adult asthma cohort. J Asthma 2021 May;58(5):586-595. [doi:
   10.1080/02770903.2020.1713150] [Medline: 31910044]
   asthma—a nationwide cohort study in Denmark. BMC Pulm Med 2015 Apr;14:15-35 [FREE Full text] [doi:
   10.1186/s12890-015-0033-z]
   wp-content/uploads/2021/05/GINA-Main-Report-2021-V2-WMS.pdf [accessed 2021-05-03]
    10.3390/ph3030514] [Medline: 27713266]
    31921830]
    31048346]
14. George M, Bender B. New insights to improve treatment adherence in asthma and COPD. Patient Prefer Adherence
15. Bârnes CB, Ulrik CS. Asthma and adherence to inhaled corticosteroids: current status and future perspectives. Respir Care
    2015 Mar;60(3):455-468 [FREE Full text] [doi: 10.4187/respcare.03200] [Medline: 25118311]
16. Mäkelä MJ, Backer V, Hedegaard M, Larsson K. Adherence to inhaled therapies, health outcomes and costs in patients
    [Medline: 23643487]
    ma exacerbations attributable to inhaled corticosteroid nonadherence. J Allergy Clin Immunol 2011 Dec;128(6):1185-1191.e2
19. Klok T, Kapteijn AA, Brand PLP. Non-adherence in children with asthma reviewed: The need for improvement of asthma
    25704083]


29. Murphy et al. JMIR FORMATIVE RESEARCH


50. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
60. Case studies. INVOLVE. URL: https://www.involve.org.uk/resources/case-studies [accessed 2021-08-10]
63. Pai A, Ostendorf HM. Treatment Adherence in Adolescents and Young Adults Affected by Chronic Illness During the Health Care Transition From Pediatric to Adult Health Care: A Literature Review. Children's Hlth Care 2011 Jan;40(1):16-33. [doi: 10.1080/02739615.2011.537934]

https://formative.jmir.org/2021/9/e28784

JMIReForm Res 2021 | vol. 5 | iss. 9 | p.18

(page number not for citation purposes)


Abbreviations

ACT: Asthma Control Test
BCT: behavior change technique
CONSORT: Consolidated Standards of Reporting Trials
GP: general practitioner
ICS: inhaled corticosteroids
MARS-A: Medication Adherence Report Scale for Asthma
mHealth: mobile health
PPI: public and patient involvement
SUS: System Usability Scale
UCSF: University of California, San Francisco
Ecological Momentary Assessment of Bipolar Disorder Symptoms and Partner Affect: Longitudinal Pilot Study

Mor Yerushalmi; Andrew Sixsmith, PhD; Ariel Pollock Star, MA; David B King, PhD; Norm O'Rourke, PhD

1Department of Psychology, Ben-Gurion University of the Negev, Be'er Sheva, Israel
2Science and Technology for Aging Research (STAR) Institute, Simon Fraser University, Vancouver, BC, Canada
3Department of Public Health, Ben-Gurion University of the Negev, Be'er Sheva, Israel
4Department of Psychology, University of British Columbia, Vancouver, BC, Canada
5Department of Public Health and Multidisciplinary Center for Research on Aging, Ben-Gurion University of the Negev, Be'er Sheva, Israel

Corresponding Author:
Norm O'Rourke, PhD
Department of Public Health and Multidisciplinary Center for Research on Aging
Ben-Gurion University of the Negev
P O Box 653
Be'er Sheva, 8410501
Israel
Phone: 972 549901808
Fax: 972 86477635
Email: ORourke@bgu.ac.il

Abstract

Background: The World Health Organization ranks bipolar disorder (BD) as the 7th leading cause of disability. Although the effects on those with BD are well described, less is reported on the impact of BD on cohabiting partners or any interactions between the two; this requires in vivo data collection measured each day over several months.

Objective: We set out to demonstrate the utility of ecological momentary assessment with BD couples measured using yoked smartphone apps. When randomly prompted over time, we assumed distinct patterns of association would emerge between BD symptoms (both depression and hypo/mania) and partner mood (positive and negative affect).

Methods: For this pilot study, we recruited an international sample of young and older adults with BD and their cohabiting partners where available. Both participants and partners downloaded separate apps onto their respective smartphones. Within self-specified “windows of general availability,” participants with BD were randomly prompted to briefly report symptoms of depression and hypo/mania (ie, BDS), positive and negative mood (ie, POMS-15; partners), and any important events of the day (both). The partner app was yoked to the participant app so that the former was prompted roughly 30 minutes after the participant with BD or the next morning if outside the partner’s specified availability.

Results: Four couples provided 312 matched BD symptom and partner mood responses over an average of 123 days (range 65-221 days). Both were GPS- and time-stamped (mean 3:11 hrs between questionnaires, SD 4:51 hrs). Total depression had a small but significant association with positive (r=–.14; P=.02) and negative partner affect (r=.15; P=.01). Yet total hypo/mania appeared to have no association with positive partner affect (r=.01; P=.87); instead, negative partner affect was significantly correlated with total hypo/mania (r=.26; P=.01). However, when we look specifically at BD factors, we see that negative partner affect is associated only with affrontive symptoms of hypo/mania (r=.38; P=.01); elation or loss of insight appears unrelated to either positive (r=.10; P=.09) or negative partner affect (r=.02; P=.71). Yet affrontive symptoms of hypo/mania were significantly correlated with negative affect, but only when couples were together (r=.41; P=.01), not when apart (r=.22; P=.12). That is, these angry interpersonal symptoms of hypo/mania appear to be experienced most negatively by spouses when couples are together.

Conclusions: These initial findings demonstrate the utility of in vivo ambulatory data collection in longitudinal mental health research. Preliminary analyses suggest different BD symptoms are associated with negative and positive partner mood. These negative effects appear greater for hypo/mania than depressive symptoms, but proximity to the person with BD is important.

(JMIR Form Res 2021;5(9):e30472) doi:10.2196/30472
KEYWORDS
bipolar disorder; couples; dyadic analyses; ecological momentary assessment; EMA; bipolar disorder; partner; relationships; mHealth; mobile apps; mental health; depression; BPD; mood

Introduction

Background
One clinical feature of bipolar disorder (BD) is variable awareness of symptoms, their severity, and impact on others; this appears especially true when manic [1]. Fortunately, smartphones today enable active and passive measurement of mood and behavior for those with mental health conditions [2-4], including BD [5-7]. Both prompted data collection [8], and embedded sensors [9] enable smartphones to capture, synthesize, and share information from those with BD and their carers (eg, spouses) [10].

For the bipolar affective disorders and older adults (BADAS) study, we randomly prompted and measured BD symptoms in the moment [11]. For this pilot study, a subset of BADAS participants downloaded the carers app onto their smartphone. We set out to demonstrate the viability of dyadic ecological momentary assessment (EMA) and to compare BD symptoms (both depression and hypo/mania) and partner mood (positive and negative affect) over time.

For this pilot study, we had two specific aims. First, to demonstrate the utility of in vivo, ambulatory assessment with BD couples (ie, yoked smartphone apps). Essentially, would persons with BD and their cohabiting carers regularly provide subjective information when randomly prompted by their respective smartphones? Assuming that ambulatory data collection proves effective, are BD symptoms and partner mood correlated when measured each day over several months? And if so, in which directions (eg, depression correlated with negative partner affect)?

Data collection via smartphone app allowed us to determine who responded first each day (ie., participant then his or her partner; or partner then participant), the interval between their respective responses, and whether they were together or apart (ie, shared vs distinct GPS coordinates).

BD Symptom Measurement
With few exceptions, BD symptom scales rely on both self-report and memory (eg, recall over the past week or month) [12]. Yet research indicates that retrospective responses are affected by recall (eg, forgetting) and various response biases [13]. For instance, end-of-day retrospective reports capture just 26% to 37% variability in mood compared to in-the-moment responses obtained earlier that day [14]. Moreover, recall accuracy declines at times of increased life stress.

This has fostered observational and objective measurement, while euthymic and symptomatic [15,16] and where people with BD work and live [17,18]. Initial research suggests that the use of smartphones can foster self-insight and help forestall BD mood episodes when patients are medication adherent [19]. This is possible because smartphones are ubiquitous today and can measure, store, and transmit data in real-time, along with location and biometric data [20-22]. This allows us to identify person-specific factors associated with the onset and maintenance of BD mood episodes [23], including the ability to sustain supportive relationships, which are important to wellness with BD over time [24].

BD Carer Well-Being
BD affects not only those diagnosed but also their family, friends, coworkers, and neighbors [25,26]. The negative impact of BD on carers includes mood episodes (depression and hypo/mania) [27], financial problems [28], and reduced social and functional well-being [29]. As a result, quality of life for BD carers can be severely impacted [30]. Compared to those caring for those with major depression, BD carers report greater burden and role strain [31].

According to Reinares et al [32], carer burden is greatest when those with BD are agitated, irritated, and depressed. Yet suicidal ideation causes carers greatest distress [33]. One challenge for BD carers is loss of control as BD mood episodes are generally unannounced, patients can present with depression, hypo/mania, or both [34], and recovery between episodes is often incomplete [35-37].

Though research examining the impact of BD on friends and family has grown in recent decades [38], all studies to date are based on retrospective questionnaire responses [31,39] or limited by very small sample sizes due, in part, to the low BD prevalence [40]. Social media recruitment for the BADAS study enabled the enrollment of an international sample of young and older adults with BD and their cohabiting spouses or partners when available.

Methods

Study App Development
The BADAS study app and data collection platform were developed, tested, and refined over 2 years, including iterative pilot testing in the field to ensure the app functioned as intended and data are reported as recorded (eg, GPS coordinates corroborated by self-reported location). Pilot testing occurred across multiple locations and time zones [8].

BADAS Study Recruitment
We first recruited 50 adults with BD living in Canada, the United States, the United Kingdom, South Africa, and Australia. Participants were recruited using microtargeted social media advertising drawn from a global population of 6.2 million English-speaking, adult Facebook users with ‘bipolar disorder interests’ (eg, members of online BD support networks). As described elsewhere in more detail [41], machine-generated algorithms calculated by social media platforms are unique not so much for their sensitivity but specificity (ie, exclusion of those who do not have BD). Thus, persons recruited via Facebook do not represent the population, but we can be

https://formative.jmir.org/2021/9/e30472

JMIR Form Res 2021 | vol. 5 | iss. 9 | e30472 | p.24
(page number not for citation purposes)
confident these are persons with BD because only persons with BD received the advertisements.

After clicking the ad, prospective participants were directed to a website describing the study; if eligible and interested, they were asked to provide their names and contact information. During screening interviews (telephone, Zoom, or Skype), prospective participants confirmed their BD diagnosis and provided emergency contact information (eg, psychiatrist). This was prudent, as bipolar disorder has the highest rate of suicide of all mental health conditions [42]. Ethics approval for this study was provided by Simon Fraser University, Burnaby, British Columbia, Canada.

**Partners of Persons with BD**

Participants were also asked if they currently lived with a spouse or partner and to provide their partner’s email address. Only with the participant’s permission did we send email requests to their respective partners, inviting their participation. Both were assured that no information would be shared between them. Despite this, most BADAS participants requested that we not invite their spouses or partners to take part. Only 3 women and 1 man agreed and downloaded the partner app onto their smartphones (eg, App Store). No partners were lost to attrition.

We purposefully recruited partners without mental health diagnoses; one couple, in which both partners had BD, was excluded. This allowed us to examine associations between normal affect and BD symptoms (ie, pathology).

**Instruments**

The bipolar disorder symptom scale (BDSx) [43] was developed for brief, ambulatory assessment of depression and hypo/mania. Respondents indicate the degree to which each of 20-mood adjectives corresponds to how they feel right now, at that moment. Research suggests a four-factor structure: two depression (cognitive and somatic) and two hypo/mania factors (elation or loss of insight and affrontive symptoms). The two depression and two hypo/mania factors are correlated, and affrontive symptoms of hypo/mania (eg, furious, disgusted, argumentative) are positively correlated with both depression factors suggesting pathways for mixed symptom presentation [44]. The construct validity of this four-factor model of symptomology was demonstrated across BD subtypes [45] and relative to quality of life with BD [46] (Figure 1).

The BDSx was developed for ecological momentary sampling of BD symptoms via smartphone app [43,44] but has also been validated for use online [45,46] and as a printed-page screening measure with BD outpatients [47,48]. In this study, $\alpha=.88$ for depressive symptoms and $\alpha=.71$ for the hypo/mania subscale.

![Figure 1. Four-factor model of bipolar disorder symptoms.](https://formative.jmir.org/2021/9/e30472)
**Profiles of Mood States**

Partner affect was assessed using the 15-item version of the profile of mood states (POMS-15), revised for daily diary research [49]. Participants are asked to rate each item on a Likert scale ranging from not at all (1) to a lot (3). POMS-15 items measure 12 negative and 3 positive emotions. This emphasis is based on research indicating that negative affect is (a) more reliably associated with individual functioning within context of acute stress and (b) more likely to be conveyed between partners and influence interpersonal processes than positive affect [48]. Internal consistency for negative POMS items measured over repeated points is high for paramedics and their spouses (α<.90) [50]. In this study, α=.87 for negative affect and α=.74 for positive affect.

**Ecological Momentary Assessment**

At recruitment, BADAS participants and partners specified “windows of general availability” in which they were randomly prompted to complete brief questionnaires on their respective smartphones. Participants were prompted twice daily to complete the BDS, (AM and PM), describe sleep quality (AM), medication adherence (AM), and any important events of the day, the importance of the event, and its impact on mood and perceived control (PM).

Participants completed a single evening questionnaire that included the POMS-15. Positive and negative affect are inversely correlated but distinct aspects of mood associated with distinct brain regions [51]. Positive and negative affect are not endpoints along a single continuum. To our knowledge, this is the first study to examine positive and negative affect as distinct constructs relative to BD symptoms, and the first to measure the effects of affrontive symptoms of hypo/mania on cohabiting carers.

BADAS participants and partners were randomly prompted up to 3 times within 30-minute data collection windows. If they did not respond within the first 20 minutes, the app prompted them again. A third and final prompt was sent 5 minutes thereafter (if they did not respond to the second prompt). Participants could select a distinct or dedicated tone to distinguish study-related prompts from other smartphone sounds [8].

The partner app was yoked to their respective participant’s app to collect couple’s data within 30 minutes. When the participant responded later than their partner’s availability, they were prompted the next morning, before noon. Both the participant and their respective partner could submit voluntary questionnaires any time if they missed a prompted questionnaire or to report a particularly salient event in the moment. Both voluntary and prompted questionnaires were time- and GPS-stamped (ie, longitude and latitude), allowing us to determine if participants and partners were together or apart when their respective questionnaires were submitted.

**Participant Remuneration**

BADAS participants were paid $1 CDN/day ($0.79 USD) if they completed both the AM and PM questionnaires when prompted. If they missed one AM or PM prompt (not both), they could later submit a voluntary questionnaire. Partners were also paid $1 CDN/day ($0.79 USD) on submission of a single PM questionnaire.

**Results**

**Viability of Ecological Momentary Assessment with BD couples**

For this pilot study, we identified 312 matched participant and partner app responses from 4 couples over an average of 123 consecutive days (mean 4 months and 3 weeks, range 65-221 days). This sample size is sufficient to detect medium to large effect sizes for correlation coefficients between BD symptomology and partner mood (where d=.80; α=.80) [52]. Although participants are few (N=4 couples), our ability to collect this volume of in vivo ambulatory data over an extended period supports our first research question (ie, N=312 matched responses). Specifically, data collection using yoked smartphone apps appears to be an effective method for long-term data collection from persons with severe mental illness and their carers (both prompted and ambient data).

**Correlational Analyses**

Our ability to collect 312 matched responses from dyads demonstrates the efficacy of ambulatory data collection with BD couples over time (mean 123 days). BADAS participants submitted the BDS before their partners completed the POMS 45% of the time (139/312; mean 4:50 hrs, SD 5:34 hrs); but most days, partners provided responses before participants (173/312, 55%; mean 1:52 hrs, SD 3:42 hrs). This sequence was largely random as it began with the participant’s PM prompt (ie, within specified PM availability). One or both responses might also have been reported voluntarily that evening, not as prompted questionnaires, which might also change the response order that day (ie, participant then partner vs. partner then participant).

This difference in completion intervals (1:52 hrs vs 4:50 hrs) reflects partners completing the questionnaire the next morning (ie, BDS submitted after partners were no longer available, following their instructions). This was not uncommon, and is consistent with the observation that those with BD are more likely to be night owls than early birds [53]. In contrast, when partners submitted the POMS first, BADAS participants also completed the BDS that evening.

We next examined correlations between total depression (cognitive and somatic symptoms), total hypo/mania (affrontive symptoms and elation or loss of insight), and partner mood (positive and negative affect). We found that depression had a small but significant association with positive (r=−.14; P=.02) and negative partner affect (r=.15; P=.01). Yet total hypo/mania appears to have no association with positive partner affect (r=−.01; P=.87); instead, negative partner affect was significantly correlated with total hypo/mania (r=.26; P=.01). This coefficient is the largest in this table, suggesting that symptoms of hypo/mania affect partners more than depression. These preliminary findings suggest that symptoms of...
hypo/mania foster sadness (ie, negative affect), not reduce positive affect (Table 1).

<table>
<thead>
<tr>
<th></th>
<th>Positive affect</th>
<th>Negative affect</th>
<th>Total depression</th>
<th>Total hypo/mania</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive affect, r (P value)</td>
<td>.b</td>
<td>-.45 (.01)</td>
<td>-.14 (.02)</td>
<td>-.01 (.87)</td>
</tr>
<tr>
<td>Negative affect, r (P value)</td>
<td>-.45 (.01)</td>
<td>-.15 (.01)</td>
<td>-.26 (.01)</td>
<td></td>
</tr>
<tr>
<td>Total depression, r (P value)</td>
<td>-.14 (.02)</td>
<td>.15 (.01)</td>
<td>-.29 (.01)</td>
<td></td>
</tr>
<tr>
<td>Total hypo/mania, r (P value)</td>
<td>-.01 (.87)</td>
<td>.26 (.01)</td>
<td>.29 (.01)</td>
<td>.b</td>
</tr>
</tbody>
</table>

aStatistically significant coefficients are in bold.
bNot applicable.

Consistent with existing research [47,48], depression and hypo/mania are positively correlated (r=.29; P=.01), suggesting that depression and hypo/mania are not inverse clinical states. Often participants reported both types of BD symptoms (eg, mixed features). By contrast, positive and negative partner affect are negatively correlated (r=−.45; P=.02).

**Partner Mood and BD Factors**

Above, we noted that the largest coefficient between BADAS participants and partners in Table 1 is between total hypo/mania and negative partner affect (r=.26; P=.01). Yet when we look more closely at BD factors, we see that negative partner affect is associated only with affrontive symptoms of hypo/mania (r=.38; P=.01). Elation or loss of insight appears related to neither positive (r=.10; P=.09) nor negative partner affect (r=.02; P=.71; Table 2).

Similarly, we noted that cognitive symptoms of depression were significantly correlated with negative partner affect (r=.18; P=.01); however, negative affect appears unrelated to somatic symptoms (r=.03; P=.58). The inverse is seen with positive partner affect, which is inversely and significantly correlated with somatic symptoms of depression (r=−.20; P=.01) but not cognitive symptoms (r=−.05; P=.43).

**Couples Together and Apart**

As previously noted, symptom and mood questionnaires were time- and GPS-stamped when submitted, allowing us to determine when questionnaires were completed and if couples were together or apart (ie, same GPS coordinates). Cognitive symptoms were significantly associated with negative affect when together and apart, and somatic symptoms were inversely associated with positive affect. Elation or loss of insight was associated with neither positive nor negative mood. What might be described as classic or quintessential mania symptoms (eg, euphoria and impulsivity) appear unrelated to partner mood when couples are together or apart.

By contrast, affrontive symptoms of hypo/mania were significantly correlated with negative affect but only when couples were together (r=.41; P=.01), not when apart (r=.22; P=.12). This result supports the construct validity of this confrontation-related grouping of symptoms. Consistent with our operational definition, these angry interpersonal symptoms of hypo/mania are experienced most negatively by spouses when couples share the same GPS coordinates. The largest coefficient in these preliminary analyses is between affrontive symptoms and negative partner affect when together (r=.41; P=.01; Table 3).
The objectives of this pilot study were to (1) demonstrate the feasibility of ambulatory data collection with BD couples and (2) identify associations between partner mood and BD symptomology over months of daily data collection. Both objectives were achieved. Moreover, preliminary analyses suggest distinct associations between depression and hypo/mania and positive and negative partner mood. GPS measurement enabled us to determine whether responses were submitted when couples were together or apart (i.e., same longitude and latitude). Though recruitment and data collection did not occur as first intended, we largely met or exceeded the standards for ambulatory assessment recommended by Trull and Ebner-Priemer [54].

BADAS participant and partner apps were yoked so that responses from both would be received within 30 minutes, fearing that between-couple effects might dissipate after more than an hour. In other words, time intervals between reporting of BD symptoms and partner mood were longer than intended. However, this makes the number and size of coefficients within couples more noteworthy. For instance, BD symptom levels reported the night before remain correlated with partner mood the next morning, suggesting that the impact of BD symptoms on partners (or partner mood on participants with BD) is not limited to minutes but appears to persist for hours maybe days. Correlation coefficients between BD symptoms and partner mood are similar to coefficients reported between partners without mental illness [50,55,56].

We examined both positive and negative partner affect in relation to BD symptomology in real-time. This proved fortuitous as we found different associations between depression and hypo/mania and positive and negative partner affect. For instance, somatic symptoms of depression are inversely associated with positive affect, whereas cognitive symptoms of depression are significantly correlated with negative affect (not positive affect).

These results are largely consistent with previous research indicating that both depression and hypo/mania affect carer well-being [33]. Our findings have the advantage of measuring both participant symptoms and partner mood each day, close in time, and over several months. Ecological momentary sampling allowed us to collect responses in real-time, unaffected by recall biases, and in familiar settings (e.g., home).

On average, we collected matched symptom-mood responses from couples each day over 3 months and 3 weeks (mean 123 days). By design, completion of app questionnaires required only 3-5 minutes. Brevity of measurement was integral to high participant retention and adherence. This high rate of participation may also be explained by participant remuneration for submission of app questionnaires. Notably, roughly 20% of BADAS participants opted to give their accumulated monies to a BD charity, suggesting both intrinsic and extrinsic motivation to participate in this study. Some participated to supplement their incomes, whereas others appeared to be motivated to contribute to BD research.

### Limitations and Future Research

The number, size, and pattern of coefficients we report warrant further study. More elaborate analyses of BD couple dynamics should be undertaken (e.g., interactions and time-lagged effects) using contemporary analyses for daily diary analyses (e.g., hierarchical linear modeling). Correlational analyses reported herein are preliminary. Most nonsignificant findings would be significant with larger samples; coefficients should be interpreted within ranges (e.g., small correlation; .20 < r < .35).

The primary limitation of this study is the sample size. We collected information from participants and partners over an extended period, but with only 4 couples; therefore, generalizability of findings is limited. This small sample size limits our ability to (ethically) report full descriptive information. Future study with more couples is needed to identify any gender or cross-national differences.

As noted above, the primary impediment to recruitment for this study was the reticence of BADAS participants to include their spouse or partner. Despite assurances that no information would be shared, the majority of participants asked that we not contact cohabiting spouses or partners. The reasons for this reluctance are not immediately apparent (i.e., we did not directly ask). Future couples research should recruit partners first, then cohabiting persons with BD, to determine if this sequence proves more effective.

As recommended by Trull and Ebner-Priemer [54], instruments used in this study were developed and validated for ambulatory assessment. For instance, the BDS$_X$ [43,44] was specifically developed to briefly measure both symptoms of depression and positive affect.
hypomania; and though we report good between-person reliability for both the BDS, and the POMS-15, ideally, we should report both within-person and between-person reliability.

Implications and Applications
Results of this study demonstrate the efficacy of EMA in dyadic mental health research. Though participants were few, we collected real-time information each day from couples over 4 months on average. Random data collection using smartphone apps is a viable methodology for longitudinal, dyadic research, including couples where one spouse lives with a chronic mental health condition. Due to the ubiquity of smartphones today, this yoked-app methodology can be applied to a range of mental health research applications. In addition, research is not limited to dyads as extended families and social networks should also be studied in vivo.

EMA data collection functioned effectively, allowing us to collect daily responses from couples when prompted. Yet allowing flexibility such as voluntary or unsolicited responses appears integral to data collection over extended periods. This, however, confounded our objective of collecting responses from both spouses within 30 minutes. Fortunately, results suggest that associations between partner mood and BD symptomology endure over extended periods (eg, the next morning). EMA research opportunities will continue to grow as mobile technology continues to advance.

EMA applications are not limited to research but also include self-care and care management. For example, push notifications (eg, SMS messages) can be generated in real-time, notifying those with BD and possibly their carers (eg, spouses) when responses suggest clinical symptomology. This can foster symptom awareness and help marshal the interpersonal resources needed to cope with and manage BD mood episodes more effectively.

Acknowledgments
Support for this study was provided by the Age-Well Network of Centres for Excellence (CRP 2015-WP6.2; NO’R) and the Canadian Institutes of Health Research, Institutes of Aging (RN 134209-259022) awarded to NO’R and A Sixsmith. Portions of this study were presented at the European Congress on Psychology in Moscow, Russia, in July 2019.

Conflicts of Interest
None declared.

References
10. O’Rourke N, King DB. Within-couple analyses of bipolar disorder symptoms and partner mood over time. 2019 Jul 04 Presented at: European Congress on Psychology; 2019, July; Moscow, Russia.


**Abbreviations**

- **BD**: bipolar disorder
- **BADAS**: bipolar affective disorders and older adults study
- **BDSx**: bipolar disorder symptom scale
- **EMA**: ecological momentary assessment
- **POMS-15**: profiles of mood states
© Mor Yerushalmi, Andrew Sixsmith, Ariel Pollock Star, David B King, Norm O'Rourke. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.09.2021. 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.
Reham Shalaby1, MD; Marianne Hrabok2, PhD; Pamela Spurvey3, CPSW; Rabab M Abou El-Magd1, PhD; Michelle Knox3, BA; Rebecca Rude3, MSW; Wesley Vuong3, MPH; Shireen Surood1, PhD; Liana Urichuk3, PhD; Mark Snaterse3, BSc; Andrew J Greenshaw1, PhD; Xin-Min Li3, MD, PhD; Vincent Israel Opoku Agyapong1, MD, PhD

1Department of Psychiatry, University of Alberta, Edmonton, AB, Canada
2Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
3Alberta Health Services, Addiction and Mental Health, Edmonton, AB, Canada

Corresponding Author:
Vincent Israel Opoku Agyapong, MD, PhD
Department of Psychiatry
University of Alberta
1E1 Walter Mackenzie Health Sciences Centre
8440 112 St NW
Edmonton, AB, T6G 2B7
Canada
Phone: 1 7807144315
Email: agyapong@ualberta.ca

Abstract

Background: Peer support is an emotional, social, and practical help provided by nonprofessionals to assist others in sustaining health behaviors. Peer support is valued in recovery-oriented models of mental health and is becoming increasingly implemented at the organizational level. Text messaging is a relatively low-cost, high-impact, and easily scalable program that uses existing technology, is devoid of geographic barriers, and is easily accessible to end users.

Objective: This study aims to evaluate the effectiveness of an innovative peer support system plus a supportive text messaging program on the recovery of discharged patients from acute psychiatric care.

Methods: This prospective, rater blinded, controlled observational study included 181 patients who were discharged from acute psychiatric care. Patients were randomized to one of four conditions: treatment as usual (follow-up care), daily supportive text messages only, peer support only, or peer support plus daily supportive text messages. A standardized self-report measure of recovery (Recovery Assessment Scale [RAS]) was completed at baseline, 6 weeks, 3 months, and 6 months. Descriptive analysis, one-way analysis of variance, and repeated measures multivariate analysis of covariance were used to examine the changes in the RAS among the study groups and over the follow-up time points.

Results: A total of 65 patients completed the assessments at each time point. For the overall sample, higher scores were found for the peer support plus text message condition compared with the text message only and treatment as usual condition on several scales (ie, willingness to ask for help and personal confidence and hope) and total score on the RAS, after 6 months of intervention.

Conclusions: Peer support plus supportive text messaging seems to result in improved recovery compared with other interventions. It may be advisable to incorporate the two interventions as part of routine practice for patients with psychiatric disorders upon hospital discharge.

(JMIR Form Res 2021;5(9):e27137) doi:10.2196/27137

KEYWORDS

peer support; recovery; controlled observational study; inpatients; mental health; supportive text messages
Introduction

Background
Peer support is emotional, social, and practical help provided by nonprofessionals to assist others in sustaining health behaviors [1]. The supporters share a similar condition as patients, successfully manage their conditions, and have received training to provide support [2]. Peer support may include activities such as advocacy, connecting resources, and experiential sharing [3]. Peer support is consistent with the recovery paradigm in mental health [4], and the purported mechanisms through which it functions [2] include knowledge sharing, modeling adaptive coping strategies, social comparison, and enhancing social support. Moreover, peer support systems can serve as an entry point into the health care system for hardly reached individuals and can provide support for those who would otherwise not engage in treatment [1]. Peer support may offer benefits to peer supporters by enhancing feelings of competence and meaning [2].

Peer support is valued in recovery-oriented models [4] of mental health and is becoming increasingly implemented organizationally [5,6]. A review reported positive outcomes, including lower inpatient service use, better relationships with providers, and increased engagement [7]. However, a rigorous evaluation of randomized controlled trials (RCTs) [8] of peer support studies reported that outcomes were mixed and often nonsignificant. In their review, the authors noted a high degree of bias and methodological limitations in the studies, including inconsistent training for peer support workers (PSWs), lack of randomization of patients, and lack of blinding of outcome assessors, and concluded that “peer support programmes should be implemented within the context of high-quality research projects wherever possible.”

The existing literature suggests that peer support is valuable, but a more rigorous methodology to evaluate peer support program outcomes is needed. This study used an RCT design to evaluate a peer support model, which incorporates, as an innovative adjunct intervention, daily supportive text messages (TxM), provision of consistent training to PSWs, adopting blindness of the assessor, and randomization of the allocated patients.

Text messaging is a relatively low-cost, high-impact, and easily scalable program that uses existing technology, is devoid of geographic barriers, and is easily accessible to end users. Several RCTs have shown significant decreases in symptomatology in psychiatric conditions after the implementation of text messaging [9,10] and high rates of satisfaction among end users [11]. During the COVID-19 pandemic, supportive text messaging has been effective in decreasing symptomatology at the general population level [12]. Incorporating such services as a standard for patients upon their discharge from acute care may significantly improve the clinical and nonclinical outcomes for these patients and the health care system.

Study Aim
The overall aim of this project is to evaluate the effectiveness of innovative peer support and supportive text messaging systems as either stand-alone or combined interventions in addition to the usual treatment for patients discharged from acute care.

Methods

Study Design
Although the initial intention was to conduct an RCT [13], subject recruitment and treatment arm allocation issues necessitated an early planned transition to a controlled observational study, as described in the following sections. Participants were recruited from June 2019 to February 2020 and were randomized into one of four conditions: (1) PSW only, (2) TxM only, (3) PSW plus TxM condition (PSW+TxM), and (4) treatment as usual (TAU). Written consent was obtained and no incentives were provided.

Initial randomization was performed by an independent statistician using the block randomization method. The generated codes were sent securely to the study coordinator to assign the recruited patients across the four arms of the study treatment groups. Participants were asked at the beginning of the interview to not reveal their treatment allocation to the researcher who would facilitate the follow-up assessments. The study database was updated by the study coordinator upon recruitment. Randomization codes were kept secured on a password-protected computer. To further maintain the blindness, the researcher conducting follow-up assessments was not granted access to the database that contained the randomization code.

The study was approved by the Health Ethics Research Board of the University of Alberta (reference number Pro00078427) and operational approval from Alberta Health Services, the regional health authority. Written informed consent was obtained from all the patients. The study was registered with ClinicalTrials.gov (trial registration: NCT03404882). In relation to the design change to a controlled observational study, the amendments to the study protocol [13] are now reflected in a revised registered trial protocol for NCT03404888.

Study Locations
The study was conducted at 5 acute psychiatric care units in Edmonton, Alberta, Canada. Patients were invited to participate in the study before their discharge.

Participants

Patients
The inclusion criteria were as follows: mental health condition (mood or psychotic disorder), imminent discharge from acute care, 18 to 65 years of age, able to provide written consent, and a mobile handset capable of receiving text messages. The exclusion criteria were as follows: inability to read the text messages from a mobile device, an addiction disorder without a mental health diagnosis, receiving PSW service before the study, or inability to commit to a sixth-month follow-up of the study.

Peer Support Workers
PSWs in this study were employed by Alberta Health Services Edmonton Zone Addiction and Mental Health Services after...
receiving 2 weeks of formal training. The PSW training program was designed by Cusick [14] and covered 13 domains: recovery and peer support; the history of recovery movement; worldview and culture; self-determination; trauma-informed care; boundaries and limits; communication and connection; the social determinants of health; impact of prejudice, discrimination, and stigma; grief and loss; crisis and recovery; goal planning; and self-care. In alignment with the literature [15], matching PSWs with our patients with respect to their baseline mental health conditions was not a criterion for assigning candidate patients to a PSW.

Demographic Characteristics of PSWs
A total of 8 dedicated PSWs were enrolled in this study (1 male and 7 females). They are employed by Alberta Health Services and occupy different positions in different health care settings within the Addiction and Mental Health portfolio. As described earlier, PSWs were not matched to our patients based on their mental health conditions, and so the mental health diagnosis of the PSWs in this study was not ascertained.

Treatment Interventions
In the PSW-only condition, a PSW met physically or virtually with the patients up to eight times over a 6-month period to offer mental health support. In the TxM-only condition, TxM were received without additional PSW intervention. In the PSW+TxM condition, participants were offered PSW services along with daily TxM. In the control arm, conventional follow-up appointments with community providers were offered but neither PSW nor TxM were provided.

1. Peer support service: patients in the PSW-only and PSW+TxM arms of the study were assigned PSWs who visited them (one to one) at the hospitals to introduce themselves and build rapport before patients were discharged into the community. PSWs visited the participants up to eight times over a 6-month period (mean 3 visits, SD 2.5). They offered the opportunity for interactive phone calls and/or texts between themselves and patients for 6 months. Phone calls or virtual meetings were offered to replace face-to-face meetings during the COVID-19 pandemic. Patients continued to receive usual community clinic or program treatments.

2. Text4Support: this is a daily supportive text message service conceived and designed by a group of psychiatrists, psychologists, mental health therapists, and patients based on cognitive behavioral therapy principles [16]. A bank of messages was generated and included different text message programs tailored for the following eight mental health domains: depression, anxiety, psychotic disorders, substance use disorders, bipolar disorder, adjustment disorders, attention-deficit or hyperactivity disorder, and general well-being. About 80% of the messages in all eight message banks had similar general well-being content; the remaining 20% targeted diagnosis-specific symptoms. Patients were enrolled by the research team to receive an assigned message bank based on their primary diagnosis by linking their phone number to the message bank through a web-based application (software). Patients in the automated TxM-only and PSW+TxM arms of this study received automated messages at 10 AM Mountain Time. Examples of these messages include the following:

- Notice the good things going on in your life right now. Often, we do not notice the good but taking a moment to do so can uplift you. (General well-being)
- When we are anxious, our thoughts often focus on future “danger.” Shift your attention to the present. What is happening right now? (Anxiety)
- Self-monitoring helps you identify and distinguish between normal changes in mood and mood swings that are problematic. (Bipolar disorder)
- Try talking quietly back to voices. Tell them they are wrong. Using the vocal part of the brain can reduce the intensity of voices. (Psychosis)

Outcomes
Participants completed measures at baseline, 6 weeks, 3 months, and 6 months. The primary outcome measure for this study was recovery, as assessed by the Recovery Assessment Scale (RAS [17]), a standardized instrument with strong psychometric properties, including high internal consistency (α=.93), test-retest reliability (r=0.88), and concurrent validity [18]. Furthermore, the scale is sensitive to changes over time [19]. This 24-item scale provides self-reported recovery ratings on a 5-point Likert scale (strongly disagree=1, disagree=2, not sure=3, agree=4, and strongly agree=5). The RAS subscales include five factors: (1) personal confidence and hope (response range 9-45); (2) willingness to ask for help (response range 4-20); (3) goal and success orientation (response range 3-15); (4) reliance on others (response range 5-25); and (5) no domination by symptoms (response 3-15). The Cronbach α coefficients for the five subscales range from 0.74 to 0.87, and the total score is positively associated with quality of life and empowerment, whereas it is inversely associated with symptoms [20]. Total scores (raw scores) were calculated for the composite RAS and for each of the five subscales and were used in the analysis of this study [19,21-23].

Sample Size
Consistent with the idea that this was a pilot study without an empirically established effect size available to aid in power and sample size calculations, the targeted sample size of 180 participants was based on existing operational resources [24].

Data Analysis
The analysis was conducted using SPSS version 20 (IBM Corp, 2011) [25]. Initially, we aimed to use intention-to-treat analysis, whereby patient data were analyzed according to their original assigned groups, regardless of the time spent in the study. However, after randomization and due to clinical logistic reasons, a significant number of patients did not receive access to the PSW service in the two intervention arms of the PSW. As stated earlier, a strategic decision was made to adapt the protocol to a controlled observational study and to change the analysis approach to as-treated, rather than intention-to-treat, to maximize the investigational value of the study without compromising or biasing outcomes.
Baseline data, including sociodemographic (age group, gender, ethnicity, education level, employment status, and relationship) and clinical characteristics (primary diagnosis and RAS five factors), were analyzed to assess between-group differences across the four arms of the study (PSW-only condition, TxM-only condition, PSW plus TxM condition, and TAU condition). The analysis was conducted using chi-square and one-factor analysis of variance (ANOVA) for categorical and continuous variables, respectively.

Age categories were generated in accordance with the quartile distribution of the age-in-years variable. RAS factors were analyzed to assess cluster differences among the four study arms across the four periods of the study, using mean and SD. A one-factor ANOVA followed by Tukey post hoc test was performed to assess the statistical differences between the study arms and corresponding mean scores on each RAS factor for all the participants who completed the follow-up assessment at any designated follow-up time point. Welch F and Games-Howell post hoc tests were performed when there was evidence of a violation of the homogeneity of variance assumption. For participants who completed assessments at all the four time points, a repeated measures multivariate analysis of covariance (MANCOVA) was used to assess the impact of the four arms of the study on participants’ scores of the RAS five factors across the three time points (6 weeks, 3 months, and 6 months follow-up), while controlling for baseline scores. With regard to MANCOVA post hoc analysis, Bonferroni corrections were used to control for multiple comparison error rate changes for post hoc pairwise analyses.

CIs and \( P \) values were used in reporting. Cases with missing values of more than one individual response per factor were excluded from the analysis. The two-tailed \( \alpha \)-level criterion for statistical significance was set at \( P \leq 0.05 \).

**Results**

**Participant Flowchart**

The study flowchart is presented in Figure 1. A total of 181 patients were recruited and randomized into four study arms (\( n=43-47 \) per condition). At 6 weeks, 64.6\% (117/181) of patients responded to the RAS survey, whereas 56.9\% (103/181) of patients responded at 3 months, and 45.9\% (83/181) of patients responded to the 6-month survey, yielding an aggregate time point response range between 45.9\% and 64.6\%.

![Study flow chart](https://formative.jmir.org/2021/9/e27137/Figure1.png)
Some patients were randomized to receive the PSW intervention with or without TxM support but did not receive PSW interventions for several reasons, including subsequent noninterest in receiving visits from PSWs and failure of PSWs to contact (Figure 1). Some of these participants continued to receive only TxM support or only TAU but attended follow-up assessments. Given the relatively small sample size of our study and the overarching objective of assessing the actual effects of the interventions, we adapted our study analysis plan to simply assess outcome data with regard to either the TAU or TxM support-only groups, reflecting the service they actually received.

### Participant Characteristics

In terms of demographic and clinical characteristics (Table 1), the overall gender balance was fairly even with 56.9% (103/181) identifying as female, 27.1% (49/181) in the range of 50-65 years of age, 69.1% (125/181) identifying as White, 55.9% (100/181) achieved a postsecondary education level, 69.4% (125/180) are unemployed, 50.8% (84/179) are single, and 50.8% (92/181) were admitted for depression and/or anxiety. Chi-square and ANOVA results indicated that participants in the four treatment arms did not significantly differ in terms of sociodemographic and clinical parameters at baseline ($\chi^2=2.7$ to $\chi^2=6.8$, $P=.08$ to $P=.91$; $F_{3,175}=0.39-1.60$, $P=.19$ to $P=.76$).

#### Table 1. Demographic and clinical characteristics of study participants in the study arms.

<table>
<thead>
<tr>
<th>Demographic and clinical variables</th>
<th>TxA, n ( %)</th>
<th>PSW, n ( %)</th>
<th>PSW+TxM, n ( %)</th>
<th>TAU, n ( %)</th>
<th>Total, n ( %)</th>
<th>Chi-square (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>18 (31)</td>
<td>14 (50)</td>
<td>18 (58.1)</td>
<td>28 (43.8)</td>
<td>78 (43.1)</td>
<td>6.8 (3)</td>
<td>.08</td>
</tr>
<tr>
<td>Female</td>
<td>40 (69)</td>
<td>14 (50)</td>
<td>13 (41.9)</td>
<td>36 (56.2)</td>
<td>103 (56.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age groups (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-30</td>
<td>17 (29.3)</td>
<td>9 (32.1)</td>
<td>8 (25.8)</td>
<td>14 (21.9)</td>
<td>48 (26.5)</td>
<td>10.9 (9)</td>
<td>.28</td>
</tr>
<tr>
<td>31-40</td>
<td>9 (15.5)</td>
<td>6 (21.4)</td>
<td>6 (19.4)</td>
<td>19 (29.7)</td>
<td>40 (22.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41-50</td>
<td>11 (19)</td>
<td>10 (35.7)</td>
<td>8 (25.8)</td>
<td>15 (23.4)</td>
<td>44 (24.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-65</td>
<td>21 (36.2)</td>
<td>3 (10.7)</td>
<td>9 (29.0)</td>
<td>16 (25.0)</td>
<td>49 (27.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.3 (6)</td>
<td>.51</td>
</tr>
<tr>
<td>Indigenous</td>
<td>5 (8.6)</td>
<td>4 (14.3)</td>
<td>4 (12.9)</td>
<td>12 (18.8)</td>
<td>25 (13.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>46 (79.3)</td>
<td>18 (64.3)</td>
<td>20 (64.5)</td>
<td>41 (64.1)</td>
<td>125 (69.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>7 (12.1)</td>
<td>6 (21.4)</td>
<td>7 (22.6)</td>
<td>11 (17.2)</td>
<td>31 (17.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educational level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.8 (6)</td>
<td>.58</td>
</tr>
<tr>
<td>Less than high school</td>
<td>6 (10.3)</td>
<td>7 (25)</td>
<td>5 (16.1)</td>
<td>11 (17.7)</td>
<td>29 (16.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school degree or equivalent</td>
<td>19 (32.8)</td>
<td>7 (25)</td>
<td>6 (19.4)</td>
<td>18 (29.0)</td>
<td>50 (27.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Above high school education</td>
<td>33 (56.9)</td>
<td>14 (50)</td>
<td>20 (64.5)</td>
<td>33 (53.2)</td>
<td>100 (55.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.7 (3)</td>
<td>.45</td>
</tr>
<tr>
<td>Employed</td>
<td>16 (27.6)</td>
<td>7 (25)</td>
<td>8 (25.8)</td>
<td>24 (38.1)</td>
<td>55 (30.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>42 (72.4)</td>
<td>21 (75)</td>
<td>23 (74.2)</td>
<td>39 (61.9)</td>
<td>125 (69.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relationship</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.1 (6)</td>
<td>.53</td>
</tr>
<tr>
<td>Married, common law, or in relations</td>
<td>11 (19)</td>
<td>11 (39.3)</td>
<td>10 (33.3)</td>
<td>16 (25.4)</td>
<td>48 (26.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>30 (51.7)</td>
<td>12 (42.9)</td>
<td>12 (40.0)</td>
<td>30 (47.6)</td>
<td>84 (46.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Divorced, separated, or widowed</td>
<td>17 (29.3)</td>
<td>5 (17.9)</td>
<td>8 (26.7)</td>
<td>17 (27.0)</td>
<td>47 (26.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Admitting diagnosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.6 (6)</td>
<td>.37</td>
</tr>
<tr>
<td>Depression and/or anxiety</td>
<td>27 (46.6)</td>
<td>14 (50)</td>
<td>16 (51.6)</td>
<td>35 (54.7)</td>
<td>92 (50.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>21 (36.2)</td>
<td>10 (35.7)</td>
<td>10 (32.3)</td>
<td>12 (18.8)</td>
<td>53 (29.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychotic disorder</td>
<td>10 (17.2)</td>
<td>4 (14.3)</td>
<td>5 (16.1)</td>
<td>17 (26.6)</td>
<td>36 (19.9)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*aTxM: supportive text messages.

bPSW: peer support worker.

cTAU: treatment as usual.
Study Outcome

For the overall sample (with variable N for each time point as shown in Table 2), ANOVA revealed a statistically significant difference between- and within-groups for scores of personal confidence and hope factor at the 3-month follow-up ($F_{3,99}=3.35; P=.02$); willingness to ask for help factor at the 6-month follow-up ($F_{3,79}=3.89; P=.01$); and total recovery score at the 6-month follow-up. Tukey post hoc tests revealed a significantly higher mean of the personal confidence and hope factor at 3 months in the PSW+TxM arm than in the TxM-only arm (mean difference 5.09, 95% CI 0.41-9.77; $P=.03$) and TAU arm (mean difference 5.09, 95% CI 0.38-9.8; $P=.03$). In addition, a significantly higher mean for willingness to ask for help was detected for the PSW+TxM arm than for the TxM-only arm (mean difference 1.87, 95% CI 0.22-3.51; $P=.02$). Similarly, for the total recovery score, the PSW+TxM arm had a significantly higher mean than the TxM-only condition (mean difference 11.78, 95% CI 3.08-20.48; $P<.01$).

For patients who completed the RAS (n=65) at all four time points (Table 3; PSW: n=13; TxM: n=19; PSW+TxM: n=12; TAU: n=19), we performed repeated measures MANCOVA, with treatment intervention as the independent variable, RAS score and subscores as the dependent variables, and baseline scores as covariates. With sphericity accepted, tests of within-subject effects indicated that neither time ($F_{10,47}=1.47; P=.18; \eta^2=0.24$) nor the interaction of time and PSW ($F_{10,47}=1.20; P=.31; \eta^2=0.20$), time and TxM ($F_{10,47}=0.48; P=.89; \eta^2=0.09$), or time and PSW+TxM ($F_{10,47}=1.24; P=.29; \eta^2=0.21$) significantly predicted RAS subscores and total scores. However, tests of between-subject effects indicated that the interaction between PSW and TxM was predictive of differences in scores on only the goal and success subscale ($F_{1,63}=4.37; P=.04; \eta^2=0.072$) and reliance on other subscales ($F_{1,63}=6.24; P=.02; \eta^2=0.10$).
Table 2. Mean and SD of the Recovery Assessment Scale total score and factor scores by study condition for patients who completed assessments at any of the four time points.

<table>
<thead>
<tr>
<th>RAS&lt;sup&gt;a&lt;/sup&gt; score and time</th>
<th>TxA&lt;sup&gt;b&lt;/sup&gt; only</th>
<th>PSW&lt;sup&gt;c&lt;/sup&gt; only</th>
<th>PSW+TxM</th>
<th>TAU&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patients, n (%)</td>
<td>Value, mean (SD)</td>
<td>Patients, n (%)</td>
<td>Value, mean (SD)</td>
</tr>
<tr>
<td>Personal confidence and hope</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (n=179)</td>
<td>58 (32.4)</td>
<td>33.4 (6.67)</td>
<td>28 (15.6)</td>
<td>34.5 (7.23)</td>
</tr>
<tr>
<td>6 weeks (n=117)</td>
<td>37 (31.6)</td>
<td>32.6 (5.91)</td>
<td>22 (18.8)</td>
<td>33.6 (5.53)</td>
</tr>
<tr>
<td>3 months (n=103)</td>
<td>33 (32)</td>
<td>31.1 (5.81)</td>
<td>16 (15.5)</td>
<td>32.4 (6.57)</td>
</tr>
<tr>
<td>6 months (n=83)</td>
<td>29 (34.9)</td>
<td>31.5 (6.37)</td>
<td>14 (16.9)</td>
<td>34.0 (5.82)</td>
</tr>
<tr>
<td>Goal and success</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (n=179)</td>
<td>58 (32.4)</td>
<td>16.5 (2.72)</td>
<td>28 (15.6)</td>
<td>16.7 (3.22)</td>
</tr>
<tr>
<td>6 weeks (n=117)</td>
<td>37 (31.6)</td>
<td>15.6 (2.44)</td>
<td>22 (18.8)</td>
<td>15.4 (3.0)</td>
</tr>
<tr>
<td>3 months (n=103)</td>
<td>33 (32)</td>
<td>15.0 (3.27)</td>
<td>16 (15.5)</td>
<td>14.9 (2.73)</td>
</tr>
<tr>
<td>6 months (n=83)</td>
<td>29 (34.9)</td>
<td>15.0 (3.24)</td>
<td>14 (16.9)</td>
<td>15.6 (2.95)</td>
</tr>
<tr>
<td>Willingness to ask for help</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (n=179)</td>
<td>58 (32.4)</td>
<td>11.5 (2.54)</td>
<td>28 (15.6)</td>
<td>11.8 (2.7)</td>
</tr>
<tr>
<td>6 weeks (n=117)</td>
<td>37 (31.6)</td>
<td>11.2 (2.52)</td>
<td>22 (18.8)</td>
<td>11.5 (1.92)</td>
</tr>
<tr>
<td>3 months (n=103)</td>
<td>33 (32)</td>
<td>11.3 (2.27)</td>
<td>16 (15.5)</td>
<td>12.0 (2.16)</td>
</tr>
<tr>
<td>6 months (n=83)</td>
<td>29 (34.9)</td>
<td>11.0 (2.6)</td>
<td>14 (16.9)</td>
<td>12.7 (1.68)</td>
</tr>
<tr>
<td>Reliance on others</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (n=179)</td>
<td>58 (32.4)</td>
<td>20.3 (3.72)</td>
<td>28 (15.6)</td>
<td>21.5 (3.29)</td>
</tr>
<tr>
<td>6 weeks (n=117)</td>
<td>37 (31.6)</td>
<td>19.4 (3.51)</td>
<td>22 (18.8)</td>
<td>20.0 (2.77)</td>
</tr>
<tr>
<td>3 months (n=103)</td>
<td>33 (32)</td>
<td>19.7 (2.88)</td>
<td>16 (15.5)</td>
<td>20.5 (2.78)</td>
</tr>
<tr>
<td>6 months (n=83)</td>
<td>29 (34.9)</td>
<td>19.5 (2.89)</td>
<td>14 (16.9)</td>
<td>21.0 (2.86)</td>
</tr>
<tr>
<td>No domination by symptoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (n=179)</td>
<td>58 (32.4)</td>
<td>9.2 (3.12)</td>
<td>28 (15.6)</td>
<td>9.7 (3.44)</td>
</tr>
<tr>
<td>6 weeks (n=117)</td>
<td>37 (31.6)</td>
<td>9.5 (2.45)</td>
<td>22 (18.8)</td>
<td>9.2 (2.74)</td>
</tr>
<tr>
<td>3 months (n=103)</td>
<td>33 (32)</td>
<td>9.5 (3.04)</td>
<td>16 (15.5)</td>
<td>9.1 (3.2)</td>
</tr>
<tr>
<td>6 months (n=83)</td>
<td>29 (34.9)</td>
<td>9.3 (3.19)</td>
<td>14 (16.9)</td>
<td>10.7 (2.09)</td>
</tr>
<tr>
<td>RAS total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (n=179)</td>
<td>58 (32.4)</td>
<td>90.9 (14.16)</td>
<td>28 (15.6)</td>
<td>94.2 (17.01)</td>
</tr>
<tr>
<td>6 weeks (n=117)</td>
<td>37 (31.6)</td>
<td>88.2 (11.6)</td>
<td>22 (18.8)</td>
<td>89.6 (13.46)</td>
</tr>
<tr>
<td>3 months (n=103)</td>
<td>33 (32)</td>
<td>86.6 (12.25)</td>
<td>16 (15.5)</td>
<td>88.9 (13.69)</td>
</tr>
<tr>
<td>6 months (n=83)</td>
<td>29 (34.9)</td>
<td>86.3 (13.88)</td>
<td>14 (16.9)</td>
<td>94.1 (13.33)</td>
</tr>
</tbody>
</table>

<sup>a</sup>RAS: Recovery Assessment Scale.<br/><sup>b</sup>TxA: supportive text messages.<br/><sup>c</sup>PSW: peer support worker.<br/><sup>d</sup>TAU: treatment as usual.
Table 3. Mean and SD of the Recovery Assessment Scale total score and factor scores by study condition for patients who completed assessments at all four time points.

<table>
<thead>
<tr>
<th>RAS * score and time</th>
<th>TxM* only (n=19), mean (SD)</th>
<th>PSW* only (n=13), mean (SD)</th>
<th>PSW+TxM (n=13), mean (SD)</th>
<th>TAU*d (n=20), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Personal confidence and hope</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>35.00 (6.63)</td>
<td>35.77 (7.50)</td>
<td>34.69 (10.37)</td>
<td>32.1 (8.23)</td>
</tr>
<tr>
<td>6 weeks</td>
<td>32.32 (6.19)</td>
<td>33.08 (5.55)</td>
<td>35.15 (6.31)</td>
<td>32.00 (6.16)</td>
</tr>
<tr>
<td>3 months</td>
<td>32.32 (5.14)</td>
<td>32.31 (7.32)</td>
<td>35.77 (5.34)</td>
<td>31.35 (7.37)</td>
</tr>
<tr>
<td>6 months</td>
<td>33.11 (6.04)</td>
<td>33.92 (6.05)</td>
<td>35.77 (3.59)</td>
<td>32.30 (6.87)</td>
</tr>
<tr>
<td><strong>Goal and success</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>16.47 (2.93)</td>
<td>16.85 (3.41)</td>
<td>16.54 (4.18)</td>
<td>15.8 (2.82)</td>
</tr>
<tr>
<td>6 weeks</td>
<td>15.26 (2.81)</td>
<td>14.85 (3.26)</td>
<td>16.85 (2.44)</td>
<td>15.55 (3.33)</td>
</tr>
<tr>
<td>3 months</td>
<td>15.32 (3.76)</td>
<td>14.77 (2.77)</td>
<td>16.54 (2.33)</td>
<td>14.70 (3.20)</td>
</tr>
<tr>
<td>6 months</td>
<td>15.11 (3.40)</td>
<td>15.62 (3.07)</td>
<td>17.38 (2.47)</td>
<td>15.80 (2.97)</td>
</tr>
<tr>
<td><strong>Willingness to ask for help</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>10.79 (3.46)</td>
<td>12.23 (3.35)</td>
<td>13.08 (1.80)</td>
<td>11.4 (3.10)</td>
</tr>
<tr>
<td>6 weeks</td>
<td>11.11 (2.71)</td>
<td>11.69 (2.06)</td>
<td>12.92 (1.38)</td>
<td>11.95 (1.93)</td>
</tr>
<tr>
<td>3 months</td>
<td>11.21 (2.72)</td>
<td>12.23 (2.24)</td>
<td>12.23 (1.92)</td>
<td>11.80 (2.57)</td>
</tr>
<tr>
<td>6 months</td>
<td>11.16 (2.65)</td>
<td>12.77 (1.74)</td>
<td>12.92 (1.98)</td>
<td>11.65 (1.76)</td>
</tr>
<tr>
<td><strong>Reliance on others</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>20.3 (3.30)</td>
<td>22.08 (3.17)</td>
<td>21.77 (.59)</td>
<td>20.25 (3.13)</td>
</tr>
<tr>
<td>6 weeks</td>
<td>19.74 (3.35)</td>
<td>20.31 (2.66)</td>
<td>21.15 (1.99)</td>
<td>20.15 (3.18)</td>
</tr>
<tr>
<td>3 months</td>
<td>19.42 (3.19)</td>
<td>20.77 (3.00)</td>
<td>21.92 (2.02)</td>
<td>20.55 (2.76)</td>
</tr>
<tr>
<td>6 months</td>
<td>19.6 (3.25)</td>
<td>21.8 (2.96)</td>
<td>21.85 (1.95)</td>
<td>20.60 (2.74)</td>
</tr>
<tr>
<td><strong>No domination by symptoms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>10.95 (3.26)</td>
<td>10.46 (3.07)</td>
<td>9.38 (3.86)</td>
<td>9.90 (2.73)</td>
</tr>
<tr>
<td>6 weeks</td>
<td>9.53 (2.86)</td>
<td>9.23 (2.95)</td>
<td>9.62 (2.36)</td>
<td>10.05 (3.02)</td>
</tr>
<tr>
<td>3 months</td>
<td>10.74 (2.54)</td>
<td>9.54 (3.31)</td>
<td>9.62 (2.82)</td>
<td>9.45 (3.38)</td>
</tr>
<tr>
<td>6 months</td>
<td>10.47 (3.10)</td>
<td>10.54 (2.07)</td>
<td>11.00 (2.12)</td>
<td>9.45 (3.20)</td>
</tr>
<tr>
<td><strong>RAS total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>93.84 (15.32)</td>
<td>97.38 (16.98)</td>
<td>95.46 (20.20)</td>
<td>89.45 (15.46)</td>
</tr>
<tr>
<td>6 weeks</td>
<td>87.95 (12.35)</td>
<td>89.15 (13.74)</td>
<td>95.69 (12.45)</td>
<td>89.70 (14.19)</td>
</tr>
<tr>
<td>3 months</td>
<td>89.00 (12.83)</td>
<td>89.62 (15.03)</td>
<td>96.08 (12.05)</td>
<td>87.85 (17.01)</td>
</tr>
<tr>
<td>6 months</td>
<td>89.47 (14.59)</td>
<td>93.92 (13.86)</td>
<td>98.92 (8.37)</td>
<td>89.55 (14.45)</td>
</tr>
</tbody>
</table>

\*RAS: Recovery Assessment Scale.
\*TxM: supportive text messages.
\*PSW: peer support worker.
\*TAU: treatment as usual.

Discussion

Principal Findings

To our knowledge, this is the first study to evaluate the effects of an innovative peer support program that incorporates supportive text messaging on recovery outcomes in patients discharged from acute psychiatric care under optimum controlled observational study conditions. An ongoing RCT in the United Kingdom is examining the effects of peer worker support for patients discharged from acute care in comparison with patients receiving TAU [26]; however, that study did not include a supportive eHealth component such as the text messaging support included in our controlled observational study.

Despite the relatively small sample size in our study, patients in the PSW+TxM group had notably higher recovery scores compared with those receiving either TxM-only or TAU. The
study measures that were included provide potentially important information regarding the mechanisms of change enacted by peer support. For example, although the mechanism of change in peer support is unclear, our results suggest that peer support may influence personal confidence and hope as well as enhance the ability of patients to ask for help.

It is notable in this study that most patients who refused to complete the follow-up assessment were in the TAU group (n=20), compared with the other groups (maximum for other groups=3). This may be explained by a lack of interest. When patients receive no actual intervention, they become less motivated to provide feedback related to the research under study. Dropout figures were the highest among the patients who were assigned to the PSW service [27]. Some patients stated that they preferred to control their path of recovery after hospital discharge. Others were not suitable candidates for this service during this initial vulnerable postdischarge period, as assessed by the PSW; many PSW-allocated patients proved hard to reach, and in such cases, PSW follow-up is usually terminated or at least significantly interrupted [27].

A recent systematic review explored different interventions, including peer support, to improve the successful transition for discharge from acute mental health inpatient care to the community [28]. In one Australian study, 38 patients achieved recovery and wellness (particularly clinical and functional recovery) after receiving peer support for 6–8 weeks, which is consistent with our results [29]. In another Australian study, 49 patients receiving peer support, as supportive packages for 8–12 hours for 1–2 weeks, reported that the intervention solidified their recovery and improved their self-confidence [30]. In contrast, in a UK study, in which 23 patients received peer support for 4 weeks and 23 were in the TAU group, there was no evidence of a significant difference between the two groups regarding hopefulness [27]. Unlike the three studies reviewed in the systematic review [28], our study findings indicate the relative impact of combined delivery of PSW+TxM compared with peer support alone, which may explain the discrepancy noted with the UK study [27].

Other studies have provided peer support to discharged patients either alone or alongside other interventions, such as environmental support or brief intervention (eg, interactive behavior change technology). However, those studies assessed outcomes other than patient recovery or reported mixed findings [31,32]. The positive effect of the combined delivery of PSW+TxM observed in our study included TxM provided to the patients that were tailored according to their diagnosis. Previous studies have reported positive benefits of receiving daily TxM in the context of mental health and addiction. For example, patients with depression alone or comorbid with alcohol use disorder reported the effectiveness of texting service on symptom recovery in terms of better management of depression and anxiety and perceived better overall mental well-being [33]. In addition, a longer time to first drink was reported after receiving TxM for 3 months and was maintained for up to 6 months [11,34]. Multiple advantages have been reported when using texting services in patients with psychotic disorders, including better medication adherence, improved clinical and functional symptoms, effective symptom monitoring, and high acceptability by end users [35-37].

Studies examining the effect of a supportive texting service for patients discharged from acute care are rare. A recent systematic review focusing on web- and mobile phone–based texting in mental health [38] reported some studies that offered texting services to patients on hospital or emergency discharge with different mental health conditions, including alcohol use disorder, bulimia nervosa, and suicide. Each of these studies reported positive outcomes, including decreased binge drinking, reduced remission rates, and achieved feasibility and acceptability by patients who attempted suicide. In contrast, our study did not produce more favorable recovery outcomes for patients who received only TxM along with TAU. This contrasts with previous findings that patients with major depressive disorder who received daily TxM in addition to usual treatment had significantly fewer depressive symptoms and improved quality of life compared with TAU [9,10]. It is interesting to note that those previous RCTs used the Beck Inventory Score changes at 3 months from baseline as the outcome measure, while this study assessed recovery outcomes using the RAS. The fact that patients in our study received text messages once daily, whereas patients in two previous RCTs [9,10] received twice daily text messages may also be related to differences in these study outcomes.

A growing body of evidence supports the paradigm of integration of health care services through multidisciplinary intervention or support. This appears to have a particularly high potential impact when patients are facing multiple and complex needs that can progress to severe forms of mental illness [39,40].

The results of this controlled observational pilot study have the potential to signal an important direction for future studies to incorporate these integration goals into peer support programs.

**Study Limitations**

Our study had several important limitations. For instance, only the RAS recovery outcome measure is reported in this study, and it is important to examine the effects of peer support and daily supportive text messaging on multiple outcomes, including quality of life, symptomatology, and health care use and functional outcomes, such as employment [13]. In addition, the RAS is a self-report outcome questionnaire and is therefore subject to social desirability and another weakness in this study. For future studies, it will be important to maximize adherence to self-reporting across the time points assessed. This can be achieved via incentives linked to completion.

Importantly, high dropout and/or nonservice provision rates for PSW among the study participants undermined the initial RCT design, thereby reducing the robustness of the study results. Consequently, to be able to access the actual impact of the interventions, we adopted a controlled observational study with a to-treat analysis rather than the original RCT plus an intention-to-treat analysis.

Although this study provided important preliminary information regarding the outcomes of peer support programs for patients discharged from acute care, the overall study sample and the individual group sizes were relatively small. Small sample sizes
reduce study power and the sensitivity of studies to detect differences between treatment groups. A multicenter study with large sample sizes will be needed to validate the results of this study and to determine the actual effect size of the various interventions forming a part of this controlled observational study.

Acknowledgments
The authors thank the unit managers and staff of the Alberta Hospital Edmonton and the psychiatric ward of the University of Alberta Hospital, the Royal Alexandra Hospital, the Gray Nuns Community Hospital, and the Misericordia Hospital in Edmonton for facilitating the recruitment of patients into the study. The authors also thank the Edmonton zone Addiction and Mental Health Peer Support Practice Council for providing peer support services for this study over and above their regular contracted work. Furthermore, the authors thank Drs Graham Gaine, Kris Uitvlugt, Julia Burbidge, and Laura Leclair as well as mental health therapists and Addiction and Mental Health service users who attend the Edmonton Mental Health Clinic located in the 108 St in downtown Edmonton for assisting in development of the cognitive behavioral therapy–informed text messages. This work was funded by the Douglas Harding Trust Fund, the Alberta Mental Health Foundation, and Alberta Health Services. The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, and approval of the manuscript; or the decision to submit the results for publication.

Authors' Contributions
VIOA conceived and designed the study, including Text4Support. RS and VIOA performed the data analysis and jointly drafted the initial manuscript with MH. RS, RMAE, and WV participated in the data collection. PS participated in the provision of peer support services, while MK and RR coordinated and supervised the peer support workers. All authors participated in the study design, reviewed and edited the initial draft of the manuscript, and approved the final draft of the manuscript before submission.

Conflicts of Interest
None declared.

References


25. IBM. SPSS Statistics for Windows, Version 20.0. IBM Corp; Armonk, NY. USA. IBM 2011:00-00.


Abbreviations
- ANOVA: analysis of variance
- MANCOVA: multivariate analysis of covariance
- PSW: peer support worker
- RAS: Recovery Assessment Scale
- RCT: randomized controlled trial
- TAU: treatment as usual
- TxM: supportive text messages

Please cite as:
URL: https://formative.jmir.org/2021/9/e27137
doi:10.2196/27137
PMID:34477565

©Reham Shalaby, Marianne Hrabok, Pamela Spurvey, Rabab M Abou El-Magd, Michelle Knox, Rebecca Rude, Wesley Vuong, Shireen Surood, Liana Urchuk, Mark Snaterse, Andrew J Greenshaw, Xin-Min Li, Vincent Israel Opoku Agyapong. Originally published in JMIR Formative Research (https://formative.jmir.org), 03.09.2021. 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: Online support groups vary widely in both goals and structures owing to the rapid development of social networking services. Several studies have shown the potential effectiveness of online support groups, such as reducing psychological distress (e.g., depression) among individuals with mental health problems. However, online support groups often do not aim at effectiveness regarding distress relief-related outcomes.

Objective: This study aims to examine whether the use frequency of online support group platform functions (U2plus) is associated with lower stigma and higher consumer activation.

Methods: A total of 350 U2plus users participated in a web-based survey. They were asked what therapy they had received in the past and how often they logged on to it, used each of its functions, and completed the following questionnaires: the Patient Health Questionnaire-9, the Devaluation-Discrimination Scale, and the General Help-Seeking Questionnaire.

Results: Regarding the therapy received, 88% (308/350) of participants had taken medication for mental health problems, and 66.6% (233/350) had received psychotherapy or mental health counseling. Regarding use frequency, 21.7% (74/341) of the participants signed in to U2plus and used its functions more than once a week. The use frequency of U2plus functions was not correlated with perceived stigma, but the use frequency of some functions was weakly correlated with help-seeking intentions from formal sources (e.g., doctors and psychologists). However, multiple regression analyses revealed that the use frequency of those functions did not uniquely predict help-seeking intentions.

Conclusions: It was suggested that online support groups may serve as an alternative treatment option for those who are already undergoing pharmacological treatment and are willing to seek help from whatever source they deem helpful.

(JMIR Form Res 2021;5(9):e21348) doi:10.2196/21348

KEYWORDS

online support group; mental health; depression; stigma; help-seeking

Introduction

Cognitive Behavioral Therapy and the Internet

Andersson [1] argues that it is essential to develop alternatives to face-to-face treatments for three reasons: (1) there are not enough experienced clinicians who are able to provide evidence-based treatment to all those who need it; (2) some people cannot access specialist clinics and general practitioners who provide and are proficient in cognitive behavioral therapy (CBT) owing to the distance from the services to where they live; and (3) some people prefer web-based over face-to-face treatments. The first reason may help reduce the stigma related to mental health problems. The development of internet-based
CBT and the rapid growth of internet access worldwide may potentially tackle these barriers to receiving psychological treatment.

After conducting a comprehensive review, Barak et al. [2] classified internet-supported interventions, which are not limited to CBT, into four categories based on their primary approaches: (1) web-based interventions, defined as “a primarily self-guided intervention program that is executed by means of a prescriptive online program operated through a website and used by consumers seeking health- and mental-health-related assistance.” Barak et al. [2] also remarked that web-based interventions attempt to create positive change or enhance knowledge, awareness, and understanding by providing sound health-related material and the use of interactive web-based components; (2) web-based counseling and therapy; (3) internet-operated therapeutic software; and (4) other web-based activities. Specifically, our study attempts to closely investigate online support groups in the fourth category.

**Alternative Effects of Online Support Groups**

People with mental illness are often interested in and willing to form connections with others through social media; specifically, young adults with mental illness were more likely to express their personal views through blogging, form friendships via social media, and connect with people via the web who share the same interests [3]. In the case of individuals with schizophrenia, despite having fewer offline relationships and less internet access, had similar tendencies to form web-based social connections with those of adults without mental illness [4]. Furthermore, online support groups, forums, and chat rooms were shown to help people to discuss their sensitive mental health conditions [5] and seek and share information related to symptoms and medications [6]. Thus, there is a bulk of the literature showing the predisposition highlighted in the first line of this section, one that could be vital during times when social distancing is required, for example, the COVID-19 pandemic has brought upon such requirements and currently persists worldwide.

After this initial phase of social networking development and group establishment (ie, people getting in touch with others and entering groups), online support groups can take on different forms. First, they have different functions; they can provide emotional, informational, and instrumental support or even (and often) a combination of all three. In particular, support groups (online) may decrease isolation, increase health information sharing, and provide role modeling [7]. Second, online support groups can have different means of access; they may be closed, open only by invitation, or even open for all. In addition, their member structure is often not fixed, so members may come and go, and many (often most) are not active; instead, they are often rather passive observers of the online discussions. Third, groups can be either self-guided by facilitators or have professionals who monitor the discussions.

Some studies have analyzed the benefits of online support groups in depth; one remarks that writing about one’s feelings and experiences associated with life challenges decreases one’s negative emotions, writing about neutral events has no such effect, and sharing one’s bad feelings with others has a tremendous relieving effect [8]. Thus, group participants may feel better when they can share their difficult emotions with others who are available to listen and show their understanding. A study conducted in a Scandinavian breast cancer online support group comprising 15 women showed that by sharing their personal stories with others in the group, participants were actively portraying their life stories and identities [9]. Similarly, a study conducted in another online support group for breast cancer patients examined the level of expression of negative emotions, showing that the possibility of conveying one’s negative feelings was imperative to ensure improvements in participants’ well-being and quality of life [10].

However, the effectiveness of online support group outcomes (eg, improvements in people’s mental health) remains largely unknown. One systematic review aimed to examine the effectiveness of online peer-to-peer support for young people with mental health problems; it identified six relevant studies to be reviewed: 3 randomized controlled trials, 2 pre-post studies, and one randomized trial [11]. These targeted studies examined various mental health–related issues, including depression and anxiety, general psychological problems, eating disorders, and substance use (ie, tobacco). Overall, two of the four randomized controlled trials yielded a positive effect for the peer-support group relative to the comparison group at postintervention: one targeting anxiety and the other targeting tobacco. The other two studies showed no evidence that a peer-to-peer support group was effective for eating disorders or depressive symptoms. Moreover, internet support groups are not risk-free; some of the risks are (1) hurtful comments could be posted, which could worsen the depression symptoms of those who read them or even encourage negative behaviors (eg, suicide attempts); (2) people could post inaccurate information; and (3) anecdotal personal stories about treatment may discourage or delay others from seeking treatment [12].

Importantly, online support groups do not always aim at effectiveness regarding distress relief outcomes [1]. Rather, they aim to promote emotional relief and an elevated sense of control. First, online support groups may challenge stigma. For example, online communities may serve as powerful venues where individuals with mental illness can challenge stigma through personal empowerment and increased hope [13]. Moreover, young adults with mental illness report that one of the primary reasons for connecting with others via the web is to feel less alone [14], because popular social media allows them to feel connected while providing a sense of relief from knowing that others share similar experiences and challenges [5]. Second, online support groups may increase consumer activation, which refers to the degree to which the individual understands that they must play an active role in managing their own health and health care, and the extent to which they feel able to fulfill that role [15]. For instance, learning from peers through web-based networks may help individuals with mental illness realize that they can make their own health care decisions [1]. In addition, previous studies have shown that when someone learns about other people’s personal experiences about facing an illness, they feel more confident in and empowered to make their own health care decisions [16]. In light of these
delineations, we aimed to test these perspectives on the functions of online support groups.

**Study Setting**

Owing to the rapid development of social networking services, online support groups take on different forms [1]. In particular, U2plus [17] is an online support group in which 23,864 users were registered (as of November 2019). It was developed to provide individuals with depression with structured peer-to-peer support while engaging in simple CBT exercises. Moreover, the name of the group has a phonetic trick to it, in that the pronunciation of U2 in English is close to that of the Japanese word *Utsu*, which means *depression*, and the term *plus* is meant to suggest that *someone else is there for you*.

After signing up, the users gain access to the psychoeducation page, from which they can learn what depression is and what keeps depression going. By applying the information that they have learned from this page, users should be able to develop a simple formulation based on a cognitive model for depression on the U2Cycle page (Figure 1). On the F2Friend page, users can present their username, sex, age group, subjective level of recovery, a brief story of their depression (eg, how they developed depression, how they have been affected, and what they want to do when they recover), and messages to other users. These messages can be viewed by other users, and they can also respond to the messages by clicking the Like button. This resource was included because previous studies suggest that reading other people’s stories leads people to feel better informed and less anxious [18,19]. Furthermore, users can engage in simple CBT techniques. On the FunCan page (Figure 2), users can post whatever they have achieved and enjoyed in their everyday life [20], to which other users can respond by clicking the Like, Great, Interesting, and Want-to-do buttons when they view this activity. On the Column page, users can post a 5-column thought record, to which other users can respond to by clicking the Empathy button when they view it.

**Figure 1.** Screenshot of the U2Cycle function.
Study Aim

This study aims to examine whether the use frequency of these U2plus functions is associated with decreased stigma, increased consumer activation, and depression levels.

Methods

Overview of the Survey

This section reports the nature of this web-based survey in accordance with the Checklist for Reporting Results of Internet E-Surveys [21]. The study was approved by the ethics committee of the International University of Health and Welfare (reference number 19-lg-46). The web-based survey was developed using Google Forms and was tested before fielding the questionnaire. In November 2019, 23,864 users registered in U2plus [17] were contacted by email and asked to voluntarily participate in a web-based survey by the end of that month (ie, before the COVID-19 outbreak). Participants assessed the survey and read the study information sheet before taking part. The survey had three screens, including the study information page, and the participants were able to review and change their answers through the back button. No incentives were offered to participants. No adaptations were made to the questionnaire, and the items were not randomized. Participants’ email addresses were collected to allow for duplicate entries from the same user to be found, but their IP addresses were not checked.

Participants

In total, we had 350 participants (350/23,864, response rate 1.46%) willing to participate, including 219 women, 118 men, and 13 who did not disclose their gender, with a mean age of 38.36 (SD 10.095) years. Regarding the therapy they received, 88.0% (308/350) had been on medication for mental health problems and 66.6% (233/350) had received psychotherapy or mental health counseling.

Measures

Overview

In addition to the standardized measures mentioned later, participants were asked about how long they had been using U2plus, to which they responded by choosing one of the five options: (1) less than 1 month, (2) 1 month to 6 months, (3) 6 months to 1 year, (4) 1 year to 3 years, and (5) over 3 years. Then they were asked how often they used each function of U2plus (ie, signing in, reacting to other users’ activities, reading
U2Friend [other users’ profile and story], U2Cycle, FunCan, and Column, to which they responded by choosing 1 of 5 options: (1) never, (2) once a year, (3) once a month, (4) once a week, (5) a few times a week, and (6) almost every day.

**Patient Health Questionnaire-9**

Patient Health Questionnaire-9 (PHQ-9) [22] was used to measure participants’ general well-being. We used the Japanese version [23]. It is a 9-item self-reported questionnaire that was originally developed to measure depressive symptoms. Items are scored on a 4-point scale (0=not at all to 3=nearly every day), with total scores ranging from 0 to 27. The questions are based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition criteria, so it can provide both a diagnosis of depression and a measurement of its severity [23]. The internal consistency of the scale ranged from 0.86 to 0.89 in the original study [22].

**Discrimination-Devaluation Scale**

The Discrimination-Devaluation (D-D) scale [24] was used to measure participants’ public stigma. We used the Japanese version [25]. It asks people how much they agree with 12 statements that begin with “Most people believe [...]”, “Most people think [...]”, or “Most people would [...]”, which are then followed by a relevant stereotype, an example of discrimination, or the opposite (ie, an accepting view or behavior). For this study, we modified the original D-D scale; in its original form, it refers to a “patient with mental health problems,” a “former patient with mental health problems,” or a person “who has been hospitalized for mental illness.” However, our adapted version referred to “a person who has received mental health treatment,” mainly because our objective was to measure perceived stigma toward a broader concept of mental health treatment, rather than stigma toward institutionalized treatment for severe mental illness. Items are scored on a 6-point Likert-type scale (0=strongly disagree to 5=strongly agree). As in the original, we constructed an index of perceived stigma by coding each response as 0, 1, 2, 3, 4, or 5, with higher numbers indicating higher perceived stigma. We then calculated the total score of the 12 items for each individual.

**General Help-Seeking Questionnaire**

The General Help-Seeking Questionnaire (GHSQ) [26] was used to measure participants’ future help-seeking intentions. The construct was measured by listing a number of potential help sources (ie, often used by people to help them deal with emotional and personal problems) and asking participants to indicate the likelihood of them seeking help from each source. Items are scored on a 7-point scale (1=extremely unlikely to seek help to 7=extremely likely to seek help). We then calculated their averages and divided them into two classifications: *Formal Sources*, computed by averaging the ratings for physicians, mental health professionals, and online counseling services, and *Informal Sources*, computed by averaging the ratings for partner, friends, and family.

**Data Analysis**

We first calculated the frequency distributions of how long the participants had used U2plus and how often they used each U2plus function. After computing the means and SDs of PHQ-9, D-D scale, help-seeking intentions from formal sources, and help-seeking intentions from informal sources, we performed bilateral correlation analyses between these scales and the use frequency of U2plus functions. Subsequently, multiple regression analyses were conducted to test if the frequency use of each U2plus function uniquely predicted the D-D scale and GHSQ. All statistical analyses were performed using IBM SPSS Statistics 24.0. Two-sided analysis was adopted, and $P<.05$ was considered as statistically significant.

**Results**

**Use Frequency for the U2plus Platform and for Each Function**

The ratios of use frequency for the U2plus platform were: 10.3% (36/350) for less than 1 month, 10.9% (38/350) for 1 month to 6 months, 16.9% (59/350) for 6 months to 1 year, 33.4% (117/350) for 1 year to 3 years, and 28.6% (100/350) for over 3 years.

Table 1 summarizes the frequencies with which they used each function in the platform. The results suggested that 9.1% (31/341) of the participants signed in to U2plus almost every day, 5.9% (20/341) signed in more than once a week, and 6.7% (23/341) signed in once a week. Thus, approximately 21.7% (74/341) of the participants used U2plus once a week or more. On the other hand, almost half of the sample used U2plus only once a year. Similar trends were observed in some U2plus functions such as Reaction and FunCan, but participants used other functions less frequently, such as U2Cycle and Column.
Table 1. Participants’ use frequency for each of the U2plus functions.

<table>
<thead>
<tr>
<th>Use frequency</th>
<th>Participant, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Never</td>
</tr>
<tr>
<td>Signing in (n=341)</td>
<td>26 (7.6)</td>
</tr>
<tr>
<td>Reaction(^a) (n=334)</td>
<td>134 (40.1)</td>
</tr>
<tr>
<td>U2Friend(^b) (n=334)</td>
<td>118 (35.3)</td>
</tr>
<tr>
<td>U2Cycle(^c) (n=333)</td>
<td>177 (53.2)</td>
</tr>
<tr>
<td>FunCan(^d) (n=335)</td>
<td>130 (38.8)</td>
</tr>
<tr>
<td>Column(^e) (n=336)</td>
<td>177 (53.2)</td>
</tr>
</tbody>
</table>

\(^a\)Reacting to other users’ activities.
\(^b\)Reading other users’ profiles and stories.
\(^c\)A cognitive behavioral therapy exercise through which users can develop a simple formulation.
\(^d\)A cognitive behavioral therapy exercise through which users can post what they had achieved or enjoyed.
\(^e\)A cognitive behavioral therapy exercise through which participants can write a 5-column thought record.

Relationships Between U2plus Function Use Frequencies, Depression, Stigma, and Help-Seeking Behavior

Table 2 presents the means and SDs of participants’ scores on each scale and their correlations with participants’ use frequency for each function of the U2plus platform. Regarding the relationships with the different scales, PHQ-9 was weakly and positively correlated with the use of the U2Cycle (P=0.02) and Column (P=0.003) functions. The Formal Sources variable of the GHSQ was weakly correlated with the use of signing in (P=0.04) and FunCan (P=0.04) functions. Finally, Table 3 presents the results of multiple regression analyses, which revealed that no use of the U2plus function uniquely predicted the D-D scale or GHSQ.

Table 2. Mean (SD) and correlations to the frequency of participants’ activity.

<table>
<thead>
<tr>
<th>Values and U2plus functions</th>
<th>Values, mean (SD)</th>
<th>Correlation factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Signing in</td>
<td>Reaction(^a)</td>
</tr>
<tr>
<td>PHQ-9(^f)</td>
<td>12.38 (6.95)</td>
<td>0.048</td>
</tr>
<tr>
<td>D-D scale</td>
<td>48.97 (10.42)</td>
<td>0.001</td>
</tr>
<tr>
<td>GHSQ-Formal(^k)</td>
<td>4.53 (1.27)</td>
<td>0.110(^l)</td>
</tr>
<tr>
<td>GHSQ-Informal(^m)</td>
<td>3.72 (1.33)</td>
<td>–0.032</td>
</tr>
</tbody>
</table>

\(^a\)Reacting to other users’ activities.
\(^b\)Reading other users’ profiles and stories.
\(^c\)A cognitive behavioral therapy exercise through which users can develop a simple formulation.
\(^d\)A cognitive behavioral therapy exercise through which users can write a 5-column thought record.
\(^e\)PHQ-9: Patient Health Questionnaire-9.
\(^f\)PHQ-9: Patient Health Questionnaire-9.
\(^g\)PHQ-9: Patient Health Questionnaire-9.
\(^h\)P=0.02.
\(^i\)D-D: Discrimination-Devaluation.
\(^j\)GHSQ: General Help-Seeking Questionnaire.
\(^k\)Help-seeking intentions from formal sources.
\(^l\)P=0.04.
\(^m\)Help-seeking intentions from informal sources.
they use it whenever they need or feel like it). Thus, for other users, the U2plus platform may be more of an on-demand service (ie, they use it whenever they need or feel like it).

As for the measured stigma, we found that stigma was not associated with any of the functions of the U2plus platform. This suggests that U2plus users may feel normalized when they discover that many people have depression, as well as when they share their experiences; however, they are constrained by the limitations imposed by the platform; for example, they can only react to other users’ posts by clicking the Like button, to prevent users from writing negative comments to each other. This is in line with two studies: one showing that loneliness and social support did not change during follow-up [28], although neither of these studies directly measured stigma, but loneliness and social support, which may affect stigma. However, they found that seeking emotional support was the most popular reason for using an internet support group [28].

We also analyzed participants’ demographic characteristics to better understand the personal characteristics of the average user of the U2plus platform, namely, those who use online support groups for depression. Participants’ average score in PHQ-9 was 12.38, suggesting that those who use the platform may be moderately depressed. Moreover, almost 88% (308/350) of the participants had been medicated for their mental health problems, suggesting that most users may be looking for something other than (or in addition to) pharmacological interventions to help them deal with their mental health issues. About 21.7% (74/341) of the participants signed in and used FunCan and Reaction more than once a week. This suggests that some users use U2plus on a regular basis, and they use simpler functions such as posting whatever they have achieved and enjoyed over the past few days, and responding to those posts, than other functions such as U2Cycle and Column, which may require more time and effort. In addition, depression measured by PHQ-9 was positively associated with the frequency use of Column and U2Cycle. Thus, for other users, the U2plus platform may be more of an on-demand service (ie, they use it whenever they need or feel like it).

## Discussion

### Principal Findings

Regarding the relationships between U2plus function use frequency, stigma, and help-seeking behavior, our results highlighted mixed outcomes: the D-D scale and help-seeking intention from informal sources were not associated with any use frequency of U2plus functions. The use frequency of the signing in and FunCan functions was correlated only with help-seeking intentions from formal sources. However, these use frequencies did not predict help-seeking intentions from formal sources in the multiple regression analysis.

We also analyzed participants’ demographic characteristics to better understand the personal characteristics of the average user of the U2plus platform, namely, those who use online support groups for depression. Participants’ average score in PHQ-9 was 12.38, suggesting that those who use the platform may be moderately depressed. Moreover, almost 88% (308/350) of the participants had been medicated for their mental health problems, suggesting that most users may be looking for something other than (or in addition to) pharmacological interventions to help them deal with their mental health issues. About 21.7% (74/341) of the participants signed in and used FunCan and Reaction more than once a week. This suggests that some users use U2plus on a regular basis, and they use simpler functions such as posting whatever they have achieved and enjoyed over the past few days, and responding to those posts, than other functions such as U2Cycle and Column, which may require more time and effort. In addition, depression measured by PHQ-9 was positively associated with the frequency use of Column and U2Cycle. Thus, for other users, the U2plus platform may be more of an on-demand service (ie, they use it whenever they need or feel like it).

### Comparison With Prior Work

Our results showed that stigma was not associated with any of the functions of the U2plus platform. This suggests that U2plus users may feel normalized when they discover that many people have depression, as well as when they share their experiences; however, they are constrained by the limitations imposed by the platform; for example, they can only react to other users’ posts by clicking the Like button, to prevent users from writing negative comments to each other. This is in line with two studies: one showing that loneliness and social support did not change during follow-up [28], although neither of these studies directly measured stigma, but loneliness and social support, which may affect stigma. However, they found that seeking emotional support was the most popular reason for using an internet support group [28].

Our results also showed that the use frequency of the signing in and FunCan functions of the U2plus platform were associated with (but did not uniquely predict) help-seeking intentions from informal sources. These results may not fully support a notion proposed by previous study, that is, when going to a medical visit, having to undergo hospitalization, or learning about others’ experiences may help individuals to feel more at ease with the situation and have a better understanding of what questions need to be asked and what to expect [1]. Many people tend to seek health information or alternative treatment options on web-based platforms after feeling dissatisfied or in clear disagreement with the advice put forth by a physician during a medical visit [29]. Confirming this statement and highlighting the positive aspects of online support groups, a study showed that peer-facilitated approaches may be important strategies for improving the quality of health care encounters, as these...
approaches help patients to find better ways to communicate with physicians, better navigate in unfamiliar health care environments, and take an active role during primary care visits [30].

Limitations and Future Directions

First, we asked all users of U2plus for their participation; this means that our participants might have had various mental health problems, not only depression. Owing to the low response rate, the sample may not have been representative of the users registered to the program in terms of basic demographic information. Thus, future studies need to recruit individuals with a specific mental health problem and establish their diagnosis by conducting structured clinical interviews. Second, we relied on participants’ self-reports about how often they used each U2plus function; thus, future studies are warranted to measure these frequencies in more objective ways, as we were not allowed to track users’ operation histories. Third, we used a cross-sectional design; thus, future studies need to use a longitudinal design and examine how these variables fluctuate with time as participants use these online support groups.

Fourth, we tested the effects of an online support group on only stigma and help-seeking intentions; therefore, future studies should examine other variables, such as sense of control, self-confidence, sense of independence, and social interaction [2]. Finally, participants were limited to those who had a good understanding of the Japanese language, and the U2plus platform was not what most would call a typical online support group regarding user interactions, as users were able to engage in some simple CBT exercises. Yet, although this is a limitation, it also raises research questions regarding what kind of online support group would be suitable for specific types of mental health problems.

Conclusions

Our findings revealed that most of the U2plus users were receiving pharmacological treatment and used U2plus as an alternative treatment option, with approximately 21.7% (74/341) of them signed in to it on a regular basis. In addition, the use frequency of some functions was correlated with help-seeking intentions from formal sources. Future studies need to more closely investigate how online support groups can help mental health stakeholders, including those who are on medication but unable to fully recover, those who are on the waiting list for evidence-based psychological treatment, and those who try to prevent relapse after recovery.

Acknowledgments

The authors would like to thank Ms M Sakuramoto and Mr K Chiba at Cotree Ltd for administrating U2plus.jp and for data collection. This work was supported by the Telecommunication Advancement Council (grant 2018HKjK).

Conflicts of Interest

None declared.

References


https://formative.jmir.org/2021/9/e21348 JMIR Form Res 2021 | vol. 5 | iss. 9 | e21348 | p.52 (page number not for citation purposes)


17. U2plus. URL: https://u2plus.jp/ [accessed 2020-09-02]


Abbreviations

- CBT: cognitive behavioral therapy
- D-D: Discrimination-Devaluation
- GHSQ: General Help-Seeking Questionnaire
- PHQ-9: Patient Health Questionnaire-9
Investigation of the Effects of an Online Support Group for Mental Health Problems on Stigma and Help-Seeking Among Japanese Adults: Cross-sectional Study

Kobori O, Yoshinaga N

URL: https://formative.jmir.org/2021/9/e21348
doi:10.2196/21348
PMID:34491206
Short Paper

Technology Acceptance Among Patients With Hemophilia in Hong Kong and Their Expectations of a Mobile Health App to Promote Self-management: Survey Study

Yin Ting Cheung¹, PhD; Pok Hong Lam², MSSc, MA; Teddy Tai-Ning Lam¹, PharmD, PhD; Henry Hon Wai Lam³, MSc (SMHS); Chi Kong Li⁴,5 MBBS, MD

¹School of Pharmacy, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, Hong Kong
²Department of Pediatrics, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, Hong Kong
³Hong Kong Haemophilia Society, Hong Kong SAR, Hong Kong
⁴Department of Paediatrics & Adolescent Medicine, Hong Kong SAR, Hong Kong
⁵Hong Kong Hub of Paediatric Excellence, The Chinese University of Hong Kong, Hong Kong SAR, Hong Kong

Corresponding Author:
Yin Ting Cheung, PhD
School of Pharmacy, Faculty of Medicine
The Chinese University of Hong Kong
Rm 801, 8th Floor, Lo Kwee-Seong Integrated Biomedical Sciences Building
Shatin, New Territories
Hong Kong SAR
Hong Kong
Phone: 852 39436833
Fax: 852 26035295
Email: yinting.cheung@cuhk.edu.hk

Abstract

Background: The lifelong management of hemophilia is demanding and complex. In July 2019, we published a review in the Journal of Medical Internet Research, summarizing telehealth interventions that facilitate monitoring of bleeding events and promoting the appropriate use of clotting factors among patients with hemophilia. This work has led to the development of a community program that aims to harness technology to promote self-management among patients with hemophilia in Hong Kong.

Objective: Before the inception of this program, we conducted a cross-sectional survey to evaluate the patients’ level of technology acceptance and identify their expectations of the use of mobile technology for self-management of hemophilia.

Methods: In total, 56 participants (75% adult patients and 25% parents of pediatric patients; 87.5% with moderate to severe disease) were recruited from a local nongovernmental organization that serves patients with hemophilia. They rated their perceived confidence and acceptance in using the new mobile technology (score 1 to 5 for each item, with a higher score indicating better acceptance) using a structured questionnaire (adapted from the Technology Acceptance Model). They also identified the top features that they perceived to be the most important components of a mobile app for the self-management of hemophilia. The Mann–Whitney U test was used to compare technology acceptance scores across subgroups of different clinical and socioeconomic characteristics.

Results: In general, the participants considered themselves skilled in using mobile apps (mean 4.3, 95% CI 4.1-4.5). They were willing to learn to use the new mobile app to organize their bleeding records (mean 4.0, 95% CI 3.7-4.3) and to manage their health (mean 4.2, 95% CI 4.1-4.5). Participants who lived in public housing (a surrogate marker for lower socioeconomic status in Hong Kong) reported lower technology acceptance than those who lived in private housing (P=.04). The most important features identified by the participants concerned documenting of infusion logs (n=49, 87.5%), bleeding events (n=48, 85.7%), and the secure delivery of the bleeding information to health care professionals (n=40, 71.4%).

Conclusions: It is encouraging to infer that patients with hemophilia in Hong Kong are receptive to the use of mobile health technology. The findings of this survey are applicable in designing the key features of a patient-centered, multimodal program harnessing mobile technology to promote self-management among patients with hemophilia. Future studies should evaluate...
participants’ acceptability and perceived usability of the mobile app via user metrics and assess clinical and humanistic outcomes of this program.

*(JMIR Form Res 2021;5(9):e27985) doi:10.2196/27985*

**KEYWORDS**

mobile health; mHealth; patients; expectations; hemophilia; chronic diseases; rare diseases; self-management

**Introduction**

Hemophilia is a rare X-linked recessive hemorrhagic disorder that affects mostly men. Hemophilia A and B are caused by deficiencies of coagulation factors VIII and IX, respectively [1]. One common severe complication of this congenital disorder is spontaneous and repetitive bleeding, particularly in the synovial joints [1]. Poor management can eventually lead to permanent joint deformity and chronic hemarthropathy, a severe type of arthritis caused by bleeding into the joints. In addition to impaired physical functioning, patients may show reduced psychosocial functioning and occupational outcomes owing to frequent hospitalization and absenteeism from work or school [2].

Currently, coagulation factor replacement therapy is the most common and effective treatment for hemophilia [3]. Patients with moderate to severe hemophilia undergo regular infusion of plasma-derived or recombinant coagulation factors to prevent spontaneous bleeding. In addition, coagulation factor replacement therapy may be administrated in the event of break-through bleeding. The introduction of home infusion therapy has also empowered patients and their families to manage the disease in a more independent manner. However, significant barriers and perceived limitations have led to a lack of adherence to treatment among patients. Patients with a poor perception of the consequences of their illness and of the necessity of treatment may reduce their adherence to prophylactic infusions [4]. Individuals with lifelong chronic conditions may present with anxiety and stress, particularly those with hemophilia, who must acquire considerable knowledge and independent management skills at a young age [5].

To address the needs of patients with hemophilia, platforms that involve various types of technology have been implemented to promote health education and good protective health behavior [6]. In July 2019, we published a review in the *Journal of Medical Internet Research* to summarize the literature on the effectiveness of telehealth interventions in improving health outcomes in patients with hemophilia [7]. This review included 16 trials and observational studies and showed that mobile technology seemed to improve patients’ adherence and accuracy in recording infusion logs and bleeding events. Studies on the provision of disease-related information and practical skills regarding the management of hemophilia yielded promising outcomes, especially among adolescent and young adult patients [7]. Patients generally reported improvements in self-efficacy in managing hemophilia after the implementation of telehealth technology, but the sustainability of the intervention depends largely on its usability and the patients’ receptivity.

Emerging studies in the literature are demonstrating the importance of understanding factors that affect the acceptance of health care technology, especially in patients with chronic diseases [8-12]. The concept of technology acceptance broadly refers to perceived usefulness or user satisfaction with a technology and often encompasses other constructs such as system usability, user feedback, perceived ease of use, attitude toward using, intention to use, and actual usage [8]. Studies have shown that affordability and accessibility are important factors influencing technology acceptance and uptake [9-12]. One Chinese study reported that individuals who were younger and had higher education attainment, higher income, and better family support were more likely to use a smartphone [11]. Other behavioral factors such as technology anxiety, resistance to change, and a lack of trust in the use of devices for self-management are associated with resistance with using mobile health apps among patients with diabetes [12]. It is important to identify these barriers during the preimplementation phase so that users’ acceptance and adoption of mobile health technology can be enhanced to maximize the success of the intervention or program.

Approximately 200 patients in Hong Kong currently have mild to severe hemophilia [13,14]. Since January 2020, a team of patient advocates, clinicians, and academic researchers has obtained sponsorship from a local philanthropic organization to establish a community program for patients with hemophilia in Hong Kong. One goal of this program is to harness mobile technology to (1) promote adherence to prophylactic infusion and recording of bleeding events, (2) facilitate timely sharing of self-documented information with clinicians to formulate or modify the treatment plans, (3) promote knowledge about self-management (eg, infusion techniques), and (4) improve social interaction in the patient community.

Before the inception of this program, we conducted a cross-sectional survey to evaluate the patients’ level of technology acceptance and identify their expectations regarding the use of mobile technology for self-management of hemophilia.

**Methods**

Between June and December 2019, participants were recruited using consecutive sampling through the Hong Kong Hemophilia Society, the only active nongovernmental organization in Hong Kong that provides services to patients with hemophilia. All participants had received a diagnosis of hemophilia A or B from a hematologist and were able to read Chinese or English. The pediatric patient (<18 years of age) surveys were completed by their parents. The Chinese University of Hong Kong Survey and Behavioral Research Ethics Committee approved this study.
before its inception (Ref SBRE-18-052), and written consent was obtained from all participants.

The participants completed a structured questionnaire that comprised 3 sections. The first section collected the participant’s demographic and socioeconomic information, and the second section evaluated his/her level of technology acceptance. Eleven questions were developed on the basis of the Technology Acceptance Model, one of the most widely applied models to describe consumer acceptability of information technology [15]. The model posits that perceived usefulness and perceived ease of use are important factors determining whether a newly introduced technology would be accepted by its potential users [8,15]. The participants rated their perceived confidence and acceptance in using the new mobile technology on a 5-point Likert scale (1=strongly disagree, 5=strongly agree). The score for each item was summed to yield a total score ranging 11-55, with a higher score indicating better acceptance of the technology. In the last section, the participants were asked to identify the top 6 features (from a list of 10) that they perceived to be the most important components of a mobile app for self-management of hemophilia. These are the 10 most common features according to a review of the existing literature on the use of mobile technology among patients with hemophilia [6,7,16-18]. The participants had the option to provide additional factors and justifications. The questionnaire was developed in traditional Chinese and was piloted with 5 patients. The questionnaire was disseminated in both paper-based and electronic formats. Self-administration of the questionnaire required approximately 10 minutes.

Descriptive statistics were used to summarize the data. The Mann–Whitney U test was used to identify differences in the technology acceptance score observed across clinically relevant subgroups: disease severity (mild to moderate versus severe), treatment type (prophylaxis versus on-demand therapy), housing type (public versus private housing, which is a surrogate marker for low versus high socioeconomic status, respectively, in Hong Kong), and highest educational attainment (for adult patients only). An exploratory analysis was conducted to evaluate the reliability (Cronbach α) of the scale. All statistical analyses were performed using SAS (version 9.4, The SAS Institute) and the tests were 2-tailed.

**Results**

In total, 56 participants completed the study (Table 1) (response rate=100%). Their average age was 37.2 (SD 14.5) years for adult patients (n=42, 75%) and 10.0 (SD 2.8) years for pediatric patients (n=14, 25%). Most of the patients (n=42, 81%) had a diagnosis of hemophilia A with a moderate to severe condition (n=49, 88%). Approximately half of the cohort (n=26, 46%) lived in public housing.
Table 1. Demographic and clinical characteristics of respondents (N=56).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Patients (n=56)</th>
<th>Parents who completed the survey on behalf of pediatric patients (n=14)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>mean (SD)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>30.4 (17.4)</td>
</tr>
<tr>
<td>Adult patients</td>
<td>42 (75.0)</td>
<td>37.2 (14.5)</td>
</tr>
<tr>
<td>Pediatric patients</td>
<td>14 (25.0)</td>
<td>10.0 (2.8)</td>
</tr>
<tr>
<td>Highest education attainment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary and below</td>
<td>30 (53.6)</td>
<td>10 (71.4)</td>
</tr>
<tr>
<td>Post-secondary and above</td>
<td>26 (46.4)</td>
<td>4 (28.6)</td>
</tr>
<tr>
<td>Types of housing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>26 (46.4)</td>
<td></td>
</tr>
<tr>
<td>Private</td>
<td>23 (41.1)</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>7 (12.5)</td>
<td></td>
</tr>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hemophilia A</td>
<td>42 (75.0)</td>
<td></td>
</tr>
<tr>
<td>Hemophilia B</td>
<td>10 (17.8)</td>
<td></td>
</tr>
<tr>
<td>Did not indicate/not sure</td>
<td>4 (7.2)</td>
<td></td>
</tr>
<tr>
<td>Treatment type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prophylaxis</td>
<td>38 (67.8)</td>
<td></td>
</tr>
<tr>
<td>On-demand therapy</td>
<td>17 (30.4)</td>
<td></td>
</tr>
<tr>
<td>Did not indicate/not sure</td>
<td>1 (1.8)</td>
<td></td>
</tr>
<tr>
<td>Disease severity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mild</td>
<td>4 (7.2)</td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td>16 (28.6)</td>
<td></td>
</tr>
<tr>
<td>Severe</td>
<td>33 (58.9)</td>
<td></td>
</tr>
<tr>
<td>Did not indicate/not sure</td>
<td>3 (5.3)</td>
<td></td>
</tr>
</tbody>
</table>

The mean technology acceptance score was 42.3 (95% CI 40.1-44.4; range 27.0-55.0). In general, the participants considered themselves skilled in using mobile apps (mean 4.3, 95% CI 4.1-4.5; Figure 1). They were willing to learn to use the new mobile app to organize their bleeding records (mean 4.0, 95% CI 3.7-4.3) and to manage their health (mean 4.2, 95% CI 4.1-4.5). Cronbach α values of the scale were .89, .90, and .87 for the overall cohort, adult patients, and parents of pediatric patients, respectively, which indicate high internal consistency.
Figure 1. Participants’ level of technology acceptance and confidence in using mobile technology (n=56). The participants reported their perceived confidence and acceptance in using the new mobile technology on a 5-point Likert scale (1=strongly disagree, 5=strongly agree). A higher score is indicative of better acceptance of the technology. Error bars represent the 95% CI values.

<table>
<thead>
<tr>
<th>Item</th>
<th>Score</th>
<th>n=56</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>The new mobile app will help me manage my health</td>
<td>3.8</td>
<td></td>
<td>3.6 – 4.2</td>
</tr>
<tr>
<td>The new mobile app will help me organize my bleeding records</td>
<td>4.0</td>
<td></td>
<td>3.7 – 4.3</td>
</tr>
<tr>
<td>The new mobile app will help me manage my treatment/infusion records</td>
<td>3.9</td>
<td></td>
<td>3.6 – 4.2</td>
</tr>
<tr>
<td>I usually find it easy to operate mobile apps</td>
<td>4.3</td>
<td></td>
<td>4.1 – 4.5</td>
</tr>
<tr>
<td>Learning how to use a new mobile app is an easy task for me</td>
<td>4.2</td>
<td></td>
<td>3.9 – 4.5</td>
</tr>
<tr>
<td>The other patients/caregivers I know are currently using mobile apps to manage their health</td>
<td>2.9</td>
<td></td>
<td>2.0 – 3.3</td>
</tr>
<tr>
<td>Using this new mobile app may change the current way of managing my health</td>
<td>3.7</td>
<td></td>
<td>3.4 – 4.1</td>
</tr>
<tr>
<td>I am not worried about privacy and leaking of my personal health data</td>
<td>3.8</td>
<td></td>
<td>3.5 – 4.2</td>
</tr>
<tr>
<td>I do not find it cumbersome to operate new mobile apps</td>
<td>3.8</td>
<td></td>
<td>3.5 – 4.1</td>
</tr>
<tr>
<td>I am willing to spend tome learning how to navigate this new mobile app</td>
<td>3.8</td>
<td></td>
<td>3.5 – 4.1</td>
</tr>
<tr>
<td>I am willing to use a mobile app to manage health if such an app is available in Hong Kong</td>
<td>3.9</td>
<td></td>
<td>3.5 – 4.3</td>
</tr>
</tbody>
</table>

Participants who lived in public housing (mean 39.9, 95% CI 37.1-42.9) reported lower technology acceptance than those who lived in private housing (mean 44.4, 95% CI 41.3-47.3; P=.04). No significant association was identified between the technology acceptance level and disease severity (P=.17), treatment type (P=.91), or education level (P=.75, adult patients only).

The most important features identified by the participants concerned documenting of infusion logs (n=49, 88%), bleeding events (n=48, 86%), and secure delivery of the bleeding information to health care professionals (n=40, 71%; Figure 2). One participant proposed health care appointment reminders to be considered as an additional feature as he struggled with “remembering his scheduled appointments with the hematology, physiotherapy, and orthopedics specialists.”

Figure 2. Participants’ preferred features of the mobile technology (n=56). Participants were asked to identify the top 6 features (from a list of 10) that they perceived to be the most important components of a mobile app for self-management of hemophilia. These features were selected on the basis of a review of the existing literature [6,7,16-18].
Discussion

Principal Findings
This survey aimed to learn about patients’ and parents’ perceptions of adopting mobile technology as an approach to manage hemophilia. We found that a large majority of our prospective users were skilled in using mobile apps. They also expressed confidence and support in learning how to use the new health apps. Participants who lived in public housing demonstrated lower acceptance and perception to the mobile technology compared to those who lived in private housing, concurrent with previous findings [8-12], which have identified lower socioeconomic status as a barrier to the acceptance and adoption of mobile health technology. These subgroups of users may require more preintervention training and regular personal contact during the implementation phase.

We applied the findings of this survey and the literature to design the key features of a patient-centered program harnessing mobile technology. The mobile app will be developed in traditional Chinese, the most common written language in Hong Kong. The first implementation phase will focus on development and promotion of the “documentation” features (ie, recording bleeding events, infusion logs, and infusion reminders). The patients and caregivers will first familiarize themselves with these primary features before expanding to secondary features that include social interaction and education functions. This staggered or waved rollout approach will allow us to identify any problems or windows of opportunity that would facilitate subsequent implementation. Considering that a telehealth intervention should not be administered alone [7,17,18], we will adopt multimodal components to complement the telehealth technology. These components include preintervention training via in-house workshops to enhance the users’ proficiency and engage patient advocates to champion this intervention. A peer-mentorship program will complement the multimedia educational platforms to enhance knowledge transfer and information utilization. To alleviate the participants’ concerns about privacy and data security, we will implement authentication systems and encryption to protect the patients’ electronic health information. The development and maintenance of the mobile app will be hosted by the information technology service center of an academic institution with well-established cybersecurity systems. In addition, the participants will be assured that only deidentified data will be exported at the back end solely for research and quality improvement purposes.

Limitations
The findings of this study have to be interpreted with caution owing to a potential selection bias; patients who were interested in the program might be more likely to have participated in this survey than those were not. The small sample size is expected because hemophilia is a rare disease. However, we recruited patients through a nongovernmental organization, and the response rate was high. This approach may have helped to establish the sampling frame and likely reduced the risk of selection bias. Other than conducting an exploratory analysis on the internal consistency of the items, we did not evaluate the other psychometric properties of the survey tool. However, the items were adapted from the Technology Acceptance Model, which is one of the most popular theoretical frameworks among similar studies conducted in other patient populations [19-21]. As this is a preintervention survey, we do expect the users’ perceptions and preference to change after we launch the mobile app. To evaluate the success of this program, our future work will include the collection of data on the participants’ acceptability and perceived usability of the mobile app on the basis of user metrics (number of downloads and installs, acquisition, stickiness, and active users). We will also assess clinical outcomes (adherence to prophylactic treatment and reduction in bleeding events) and humanistic outcomes (users’ satisfaction and health-related quality of life).

Conclusions
Initiation of this multimodal program that includes mobile technology would be parallel with the vision embodied by “The Smart City Blueprint for Hong Kong” [22], which was introduced by the Government of the Hong Kong SAR to facilitate the use of innovation and technology to improve people’s quality of life. It is encouraging to infer that patients with hemophilia in Hong Kong are receptive to the use of mobile technology as part of a multimodal program to improve self-management of their health. However, like with any service project, there is a risk of failure if the program is not implemented in a thoughtful way. As identified by the study participants, there is a need for ongoing promotion and monitoring of usability, particularly in the early implementation phases. Additionally, continual maintenance, quality assurance, and data security should be ensured to maximize the sustainability of this project.

Acknowledgments
We would like to acknowledge members of the Hong Kong Haemophilia Society for participating in this study, as well as Mr George Yuen (Cloud Computing Division, Information Technology Services Centre, The Chinese University of Hong Kong) and his team for developing the mobile app. This work is partially funded by the Jockey Club Charities Trust for the project “Eyes on Haemo”—a self-management multi-modal care program (S/N 2020-0106)—and the Chinese University of Hong Kong Research Incentive Scheme—Top-up for General Research Fund Projects (Ref: 4442660).

Conflicts of Interest
None declared.
References


22. Hong Kong Smart City Blueprint. URL: https://www.smartcity.gov.hk/ [accessed 2021-02-09]
Using the Think-Aloud Method to Assess the Feasibility and Acceptability of Network Canvas Among Black Men Who Have Sex With Men and Transgender Persons: Qualitative Analysis

Natalie D Crawford¹, PhD; Dorie Josma¹, MPH; Kristin R V Harrington², BS; Joseph Morris¹,³, MPH; Alvan Quamina⁴, PhD; Michelle Birkett⁵, PhD; Gregory Phillips II⁵, PhD

¹Department of Behavioral, Social, and Health Education Sciences, Rollins School of Public Health, Emory University, Atlanta, GA, United States
²Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, United States
³Department of Population Health, School of Public Health, Georgia State University, Atlanta, GA, United States
⁴NAESM Inc, Atlanta, GA, United States
⁵Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Corresponding Author:
Natalie D Crawford, PhD
Department of Behavioral, Social, and Health Education Sciences
Rollins School of Public Health
Emory University
Grace Crum Rollins Building
1518 Clifton Rd
Atlanta, GA, 30322
United States
Phone: 1 404 712 9445
Email: ndcrawford@emory.edu

Abstract

Background: Characteristics of an individual’s social network have been important factors in understanding infectious disease transmission patterns. Social network data collection is generally time and resource intensive, yet it is crucial to our understanding of the complex epidemiologic landscape of human behaviors among stigmatized social groups.

Objective: We sought to evaluate the feasibility and acceptability of a self-administered social network data collection tool, Network Canvas, among Black men who have sex with men (BMSM) and transgender persons using the think-aloud method, which is a robust and flexible research technique used to perform usability testing.

Methods: We piloted a self-administered network interview within the Network Canvas Software Suite. Participants aged 18 years and older were recruited through a community-based organization in Atlanta, GA, and were included based upon their willingness to share information on sexual behaviors and drug use for themselves and their social networks. A semistructured interview guide was used to document cognitive decision-making processes while using the tool. Recorded interviews were transcribed verbatim, and thematic analyses were performed.

Results: Among 7 BMSM and transgender participants, three main themes were identified from cognitive processes: (1) the utility, (2) navigation, and (3) intuitive design of Network Canvas. Overall, Network Canvas was described as “easy to use,” with suggestions mainly directed toward improving navigation tools and implementing an initial tutorial on the program prior to use. Participants were willing to use Network Canvas to document their social networks and characteristics. In general, observed verbal responses from participants matched their behavior, although there were some discrepancies between verbal affirmations of use and understanding versus external observation.

Conclusions: We found Network Canvas to be a useful new tool to capture social network data. Self-administration allowed participants the opportunity to provide sensitive information about themselves and their social networks. Furthermore, automated name generation and visualization of an individuals’ social network in the app has the potential to reduce cognitive burden during data collection. More efficient methods of social network data collection have the potential to provide epidemiologic information to guide prevention efforts for populations with stigmatized health conditions or behaviors.
Introduction

Social networks are understood as patterns of stable interactions among people [1,2] that can be categorized as instrumental, supportive, disruptive, burdensome, or neutral. Infectious disease transmission, such as HIV transmission, requires interactions between at least two individuals. Thus, characteristics of an individual’s social network have been important factors in understanding infectious disease transmission patterns. Indeed, various social network characteristics have been identified as robust predictors of HIV transmission. For example, an individual’s network size, demographics of the network, and the individual’s position in their own network are highly linked to the risk of sexually transmitted infections [3], and sexual behaviors, and health outcomes [11]. Although the use of paper social network inventories has been collected using standard data collection methods that ask research participants to list people within their networks during a specified period and then to provide potentially extensive information about each alter’s demographics, perceived behaviors, and health outcomes [11]. The ability to capture valid reports of social network data is crucial in order to assess the relationships between social networks and HIV. Social network inventories have mainly been collected using standard data collection methods that ask research participants to list people within their networks during a specified period and then to provide potentially extensive information about each alter’s demographics, perceived behaviors, and health outcomes [11]. Although the use of paper social network inventories is effective and has produced reliable reports even with historical accounts of social networks, the data collection, entry and cleaning processes for these data are cumbersome for both participants and researchers [11]. Moreover, the fatigue related to this method of social network data collection may lead to underreporting of novel survey measures. To our knowledge, this method has not been used in health research to understand whether study participants can adequately self-administer an electronic social network data collection platform by understanding the natural flow of information of their cognitive process while completing a task [22,23]. The think-aloud method is widely used in the disciplines of psychology [23], engineering [24,25], education [22,26], and public health [27,28]. In public health, think-aloud methods have been used to assess participants’ cognitive understanding of novel survey measures. To our knowledge, this method has not been used in health research to understand whether study participants can adequately self-administer an electronic social network data collection tool by understanding the natural flow and expectations of the information being requested by the program. In this article, we utilized the think-aloud method to understand BMSM and transgender persons’ cognitive processes and then assessed the feasibility and acceptability of Network Canvas for personal social network data collection.

Methods

Design

Northwestern University partnered with researchers at the Rollins School of Public Health at Emory University to pilot the Network Canvas software app. In order to assess the feasibility and acceptability of using Network Canvas among BMSM and transgender persons, a semistructured interview guide was used to instruct participants in how to verbalize their cognitive decision-making process in real-time while completing a task or using a tool on Network Canvas. All study activities were approved by the Emory University Institutional Review Board.
Recruitment

A total of 7 participants (N=7) were typically purposively recruited in collaboration with a community-based organization in Atlanta, GA. To be included in the study, all participants had to be aged 18 years or older and willing to share personal information on sexual behaviors and drug use for themselves and their social networks. Participant ages ranged between 19 and 28 years old. Participants self-identified as BMSM (n=5), transmen (n=1), and transwomen (n=1).

Data Collection and Analysis

Data for this study were collected as a part of a parent study in which participants were interviewed in order to evaluate their willingness to receive sexual health services in pharmacies. Participants were subsequently asked about their willingness to participate in the feasibility and acceptability testing of Network Canvas. Prior to data collection, researchers downloaded the Network Canvas Architect Application, the component of the Network Canvas Software Suite that deploys study protocols to a password-protected iPad. Architect software (version 4.0.0; Complex Data Collective) was used.

In the Network Canvas protocol, participants were asked to identify their social networks and engagements with these networks in the last 3 months (see Figure 1 for example screenshots). In the name generator, participants assigned fictitious names to people in their social networks; these names were then linked to their reported demographic information (ie, age and sex), sexual health status, and drug use behaviors (ie, cocaine, heroin, or opioids). Participants were asked questions such as, “Who did you get together to hang out with or socialize?” “Who did you have sex with?” and “Who did you use drugs with?” Participants were also instructed to place each network on a sociogram. The names of individuals who were closest socially to the participants were arranged in the center of the circle, and those who were the least socially close to participants were placed on the margins of the circle.

Researchers used think-aloud methods in order to guide participants to verbally describe their cognitive process by using the features and tools while completing the Network Canvas social network inventory. Participants were asked to actively describe their understanding of the software program as they navigated it to complete the survey questions. Following the interview, participants were asked questions about their experience using the software such as, “Did you find the application intuitive?” and “How was your experience using the Network Canvas application to answer these questions?” All interviews were recorded using an audio-recorder. Researchers took observational notes to capture nonverbal cues not captured on the audio recorder. All data were uploaded to a secure computer server to maintain participants’ confidentiality. Each participant received a US $50 gift card to compensate for their time upon interview completion.

Research assistants with a masters-level training in qualitative methods transcribed the audio-recorded interviews verbatim. Participants’ verbatim transcripts were complemented with their respective observational notes to enhance the analysis of both verbal and nonverbal cues. NVivo 12 (Version 12; QSR International), a qualitative data analysis software, was used to perform a thematic analysis of the participants’ interviews and observation notes. Three transcripts were analyzed and used to develop a codebook with inductive codes, definitions, and in-text examples. The codebook was then used to code the remaining transcripts and generate findings. Saturation was determined when new emergent themes were no longer present in the data. Once saturation was reached, participant recruitment was stopped.
Results

Cognitive processes of participants were captured while navigating through the Network Canvas application, and close attention was paid towards participants’ processes using the application’s features and tools. Analyzed data were organized into three main themes: (1) the utility, (2) navigation, and (3) intuitive design of Network Canvas.

Utility of Network Canvas

All participants described Network Canvas as “easy to use.” They also shared that features within the application were useful in collecting and organizing personal network information. Specifically, one participant explained that the name generator was useful in allowing names to be added and populated on the sociogram for later use.
It was easy. It was very easy. [Participant #107] It’s an easy process and you don’t have to add them [names of social networks members] again. Just a simple tap. [Participant #112]

Navigation of Network Canvas
Although participants responded positively to the utility of the application, most of them expressed ways in which to improve navigation. The name generator and sociogram required participants to perform different data input tasks either by clicking, tapping, or dragging. Participants (67/85.7%) found that these various tools between pages made it harder to engage with the application. During the interviews, participants asked the interviewers how to navigate between questions and pages.

So, it’s asking me a question but how am I supposed to—click this button here to answer…So, I guess I hit this arrow to move forward?... So, you’re saying if they’re highlighted, it means selected? [Participant #108] I’m going to the next question. Do I just click on them?... Okay so this person name is here, how do I get to the other person name? [Participant #109] The tapping and dropping and dragging…that’s a lot...Oh! There we go. Untap. So, you said...Yeah, no, that wasn’t clear. [Participant #110]

To help participants better navigate the application, researchers added navigation instructions at the beginning of each section. However, interviewers observed that most participants proceeded to the survey questions without reading the navigation instructions. One participant openly shared his thoughts about the writing prompts.

It was like a lot of writing [haha] not going to lie. [Participant #112]

Intuitive Design of Network Canvas
Overall, all participants responded positively to the application’s design and felt that the application was generally intuitive. One participant expressed that he would be able to use the application on his own without any assistance. The remaining participants expressed that as they moved through the survey, they used the features and tools appropriately.

Once you get the hang of it it’s easy. [Participant #110]

Several participants offered recommendations on ways to improve the design of the application. One participant mentioned that a “down arrow” tool be changed to “Next.” He expressed that the proposed tool would make navigating through the application more intuitive.

So maybe like the down arrow—instead of it being a down arrow, maybe it could say “next” or something like that because I’m still looking for the “next” button in the back of my mind. I’m just used to seeing next. [Participant #108]

One participant was repeatedly shown how to use many of the tools for the name generator and sociogram features. Although he rarely verbalized his concerns, the interview observed that he faced challenges navigating through the application. However, after being shown how to use the tools, he felt confident and found the application easy to navigate. He expressed that people using the application should be taught how to navigate through it.

I believe someone should be there to instruct the person before they use it, and then let them go off on their own after they’ve instructed them on how to use it and, what to do, and how to answer the questions. [Participant #107]

Discussion
Principal Results
In this study, we used the think-aloud method to assess the feasibility and acceptability of an electronic social network data collection tool, Network Canvas, among BMSM and transgender persons. The think-aloud method was used to assess participants’ cognitive process while completing the social network inventory in Network Canvas. Researchers followed the participants’ cognitive process to understand whether it matched with what was expected of each participant. While in most cases participants’ verbal responses matched their behavior, there were some discrepancies between their verbal affirmation of their use and understanding of the feature versus external observation. Our results suggest that participants were willing to use Network Canvas and found it to be feasible and generally easy to use. However, the sociogram feature and some of the navigation tools required the most instructions for participants. Although participants believed that the design of Network Canvas was easy to understand, they had suggestions for improvement, including more intuitive forward buttons with labels noting the next step. They suggested the inclusion of a brief tutorial before allowing participants to complete the social network inventory on their own. They also noted a need for features and tools to be consistent on each data collection page to improve the application’s intuitiveness.

Comparison With Prior Work
Instead of interviewer-led assessments, self-report procedures have been utilized in research on vulnerable populations with stigmatized health conditions or behaviors due to their ability to collect valid and reliable measurements of risk behaviors [29,30]. Previous studies have shown that when self-reporting procedures are structured in a way to maximize response accuracy, valid assessments can be collected [31,32]. In this study, navigation prompts were not highly utilized and may not be an effective way to communicate instructions. Participants suggested the implementation of both an orientation prior to using the application, as well as more informative navigation buttons, which are both potential methods to gather more detailed and accurate information in future studies.

The collection of accurate social network information from at-risk populations can be difficult and resource intensive, yet it has the potential to inform targeted interventions with greater impact [33]. Thus, it is crucial to further develop more feasible and efficient methods to collect social network data that can be implemented in a wider range of settings. For example, social
network name generation through either paper or interviewer-based methods places a significant cognitive burden on the participant, and it is susceptible to interviewer effects if prompts are asked differently each time [34,35]. This challenge highlights the utility of the Network Canvas tool, which allows for interactive building and complete visualization of the participant’s social network to reduce cognitive burden and is not sensitive to interviewer effects due to the standardized application platform. Previous research has used the think-aloud method to develop and adapt measurement scales as well as websites and other electronic applications [36]. For example, one study used think-aloud methods to improve the comprehensibility of pediatric antiretroviral therapy adherence measurement items to adapt surveys to cultural context [37]. Another study used think-aloud methods to assess the usability of a smartphone app for the purposes of helping people reduce their alcohol consumption [38]. The findings of this study have been used to further the development of the Network Canvas software. Although interviewer-assisted data collection using Network Canvas is the most optimal, this preliminary data establishes evidence that self-administration of this program is possible, particularly with short, visual tutorials to orient the participant.

Limitations
This study has a number of important limitations. First, due to the small sample size of this study, we are unable to generalize our findings nor generate greater user feedback data to further improve the usability of Network Canvas. However, during data collection, we reached saturation among participants. Second, there is a risk of reporting bias, in which participants may have given answers in the direction they perceived to be expected by the researchers. Furthermore, participants may have only selected to verbalize the thought processes they wished to share with interviewers (ie, social desirability bias). Third, there is also a risk of acquiescence bias, if participants’ responses tended to have more positive connotations or associations within their stated social networks. Fourth, although we used the think-aloud method to understand how the participant cognitively processed the social network inventory software, this method does require that participants speak aloud throughout the interview, and this was not performed consistently throughout the interview nor across participants, thus likely limiting the observational data we were able to collect [23]. To mitigate the potential effects of this, research assistants were trained to remind and incite verbal feedback from participants, as well as collect observational data that were used to supplement participants’ verbatim transcripts during data analysis. Fifth, the functionality of the name generator within the Network Canvas software in which fictitious names were assigned to people in one’s network may limit the generalizability of these results to future studies that may require the use of the real names of contacts in a network. Lastly, although this was part of a larger study focused on assessing BMSM’s willingness to receive sexual health services in pharmacies, the research and data collection protocol was written solely for the purposes of this study.

Conclusions
Overall, this study showed that Network Canvas is a useful tool to capture social network data, and it has the potential to be a widespread, efficient data collection method. Its ability to be self-administered allowed for participants to provide confidential data about their social networks and their engagements with those social networks. Through observational data, participants asked questions on how to use tools within Network Canvas. Therefore, it would be useful to include short tutorials to enhance participants’ ability to navigate through the application. This study lays the groundwork for further research to assess usability and feasibility in a larger sample of people from different cultural backgrounds who may not be as familiar with the technology. Further evaluation of this self-administered software application to collect social network data has the potential to provide rich descriptive epidemiologic information that can help to guide future prevention strategies.

Acknowledgments
We would like to thank the participants for their time and engagement in this study. We would also like to acknowledge the National Institute of Mental Health (NIMH) for its support under award number R34MH119007, and the National Institute on Drug Abuse (NIDA) for its support under award numbers R01DA042711 and K08DA037825.

Conflicts of Interest
All authors have no relevant disclosures.

References


Abbreviations

BMSM: Black men who have sex with men
MSM: men who have sex with men
NIDA: National Institute on Drug Abuse
NIMH: National Institute of Mental Health
Collecting Social Media Information in a Substance Use Intervention Trial With Adolescent Girls With Lifetime Substance Use History: Observational Study

Lili M C Ramos¹, BA; Joseline Delgadillo², BSW; Sarah Vélez², BA; Emily Dauria², PhD, MPH; Jamie Salas², MSc, LMFT; Marina Tolou-Shams², PhD
¹Department of Psychology, John Jay College of Criminal Justice and the Graduate Center, City University of New York, New York, NY, United States
²Department of Psychiatry and Behavioral Sciences, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Marina Tolou-Shams, PhD
Department of Psychiatry and Behavioral Sciences
University of California, San Francisco
1001 Potrero Ave
San Francisco, CA, 94110
United States
Phone: 1 (628) 206 2212
Email: Marina.Tolou-Shams@ucsf.edu

Abstract

Background: Adolescents with juvenile legal system contact face numerous barriers to participation in behavioral health intervention research, including housing disruption, legal privacy concerns, and systems mistrust. Technology, such as social media, may be a novel and developmentally appropriate adolescent research study engagement and retention tool.

Objective: We examined data on social media information collected for study retention purposes from adolescents participating in a substance use intervention trial.

Methods: Data were collected as part of a randomized controlled trial determining efficacy of a group-based substance use intervention for girls and young women (12-24 years) with substance use histories referred from legal and school systems in the United States. Baseline demographic and social media information was analyzed from the subset of 114 adolescent girls (mean age 15.7 years; range 13-18 years), of whom 31.6% (36/114) were legally involved, 87.7% (100/114) belonged to minoritized racial/ethnic groups, and 32.5% (37/114) received public assistance.

Results: Most girls (74/114, 64.9%) provided at least one social media account (Instagram, 95% [70/74]; Facebook, 27% [20/74]; and Twitter, 11% [8/74]) during study enrollment. Legally involved girls were significantly less likely to provide social media information than school-referred girls (44% [16/36] versus 74% [58/78]; \( \chi^2 \) \[N=114\]=9.68, \( P=.002 \)).

Conclusions: Obtaining social media information for study retention purposes from adolescent girls with lifetime substance use appears possible; however, particular subgroups (ie, legally involved girls) may be less likely to provide accounts. Factors shaping legally involved girls’ willingness to provide social media information, including mistrust and privacy concerns, and the impact of researcher’s access to social media information on study retention are critical directions for future research.

Trial Registration: ClinicalTrials.gov NCT02293057; https://clinicaltrials.gov/ct2/show/NCT02293057

(JMIR Form Res 2021;5(9):e25405) doi:10.2196/25405

KEYWORDS
adolescent girls; legal involvement; substance use; social media; health intervention

Introduction

Large, representative samples in substance use intervention research are essential to best inform substance use treatment delivery policy and practices with adolescents with juvenile legal system contact (herein referred to as “legally involved adolescents”) and across the behavioral health cascade of care [1]. Prior research with legally involved adolescent populations has identified challenges to engaging and retaining participants...
in longitudinal intervention trials, suggesting a need for novel, age-specific contact methods [2,3]. High rates of social media use among adolescents in the United States are leading researchers to explore social media as a tool for recruiting, retaining, and intervening with adolescent research participants (eg, [4-6]), particularly those who may be harder to reach using traditional retention strategies (eg, in-person, voice calls). Social media allows adolescents to control how they present themselves to their social networks at a developmental stage when self-esteem, self-exploration, autonomy, and identity development matter significantly [7]. As adolescents rely increasingly on social media to fulfill these needs, their elevated time spent on the platforms suggests that social media could be a developmentally appropriate engagement strategy for researchers to reach and connect with adolescents on the platforms they already use to connect with peers, develop their individual identities, and carry out their day-to-day social lives [8].

Social media is also an efficient mechanism to recruit diverse (ie, in race/ethnicity, gender, socioeconomic status) adolescent groups into health research [6,9,10]. Social media may be especially useful for substance use intervention trials with legally involved adolescents, for whom researchers may face study recruitment and retention obstacles, such as disproportionate housing disruption that precludes consistent in-person contact [3,11] and high rates of household poverty that can result in frequent cellphone service disruption [12]. Social media platforms present a possible solution because they are accessible across multiple devices if phones are lost or stolen and via internet connections if phone plans are disrupted. Participants’ willingness to provide social media information for study communication must be examined, particularly among minoritized adolescent populations who may have differential willingness to provide social media information. Specifically, legally involved adolescents may be reluctant to provide social media information due to privacy concerns. Prior work has described how law enforcement agencies use social media for surveillance purposes, such as investigating and monitoring individuals’ activities online [13,14]. Adolescents who use substances may be especially reluctant to provide social media information due to concerns that images posted online might capture substance use and could lead to further system involvement. Surveillance by law enforcement has historically impacted People of Color disproportionately, which may also contribute to adolescents’ decisions to share social media information [13]. Empirical literature on using social media for study retention purposes to date has not included legally involved adolescent substance use research populations and has primarily focused on evaluating only 1 social media platform (ie, Facebook) [15,16].

As part of a larger substance use intervention trial with adolescent girls (legally involved and at-risk for legal involvement by virtue of substance use history), we examined data on social media account information provided to research staff for retention efforts. We hypothesized that legally involved girls would be less likely to provide social media information for study contact purposes than girls at-risk for legal involvement who were referred from schools (hereafter “school-referred girls”).

**Methods**

This study includes baseline adolescent demographic and social media data collected between 2016 and 2019 as part of a federally funded, randomized controlled trial (Project VOICES) testing the efficacy of a gender-responsive, group-based substance use intervention [17] with girls and young women (12-24 years). To examine adolescent-specific patterns, we analyzed data from self-identified female adolescents (hereafter “girls”), aged 13-18 years (114/132, 86.4%) who could legally hold social media accounts according to the Children’s Online Privacy Protection Act (ie, age ≥13 years). Eligible girls had to endorse lifetime history of substance use (alcohol, cannabis, or other drugs) on a private, computerized screener. Girls were ineligible for participation if their substance use treatment need required a higher level of care than an outpatient, if they chose to participate in an alternate substance use intervention, or had observable cognitive or developmental delays or active psychosis that would preclude group participation. Girls were recruited from juvenile probation and diversion and public school settings in a large metropolitan area on the West Coast of the United States. All girls referred from probation or diversion programs were living in the community (and not detained); school counselors referred girls from the schools who they determined might benefit from substance use intervention participation.

Eligible girls completed study assent, consent, and locator information for study follow-up. A Certificate of Confidentiality was obtained from the National Institutes of Health as an additional protection for participants’ privacy and was reviewed with adolescents prior to enrollment. For locator purposes, adolescents were asked to provide (1) cellphone numbers; (2) social media account information (Facebook, Instagram, or Twitter); (3) home address; and (4) contact information for at least three individuals who could help locate the adolescent if other contact methods were unsuccessful. Demographic data (eg, age, race, ethnicity) were collected as part of a private, computerized baseline assessment. Following the baseline assessment, girls were randomized to either an active (VOICES) or control (GIRLHealth; psychoeducational health curriculum) intervention, each consisting of twelve 60-minute group sessions (6-8 girls/group) (see [18] for a detailed description of full trial methods). Participants completed study assessments pre-(baseline)-, mid-, and immediate postintervention completion, and at 3 and 6 months after the intervention. All study procedures were approved by the Institutional Review Board of the Principal Investigator’s institution (MT-S; University of California, San Francisco).

**Results**

Approximately one-third of girls were legally involved (36/114, 31.6%; Table 1). Girls were on average 16 years, predominantly belonged to minoritized racial and ethnic groups (12.3% [14/114] White, non-Hispanic/Latino), and 32.5% (37/114) reported receipt of public assistance (eg, food stamps). There

---

**Table 1.**

<table>
<thead>
<tr>
<th>Group</th>
<th>Number (n)</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>114</td>
<td></td>
</tr>
<tr>
<td>Legally involved girls</td>
<td>36</td>
<td>31.6%</td>
</tr>
</tbody>
</table>


were no age (t_{112}=0.847, P=.40) or race/ethnicity ($\chi^2_{4} [N=114]=4.92, P=.30$) differences between legally involved and school-referred girls. School-referred girls were less likely to report receipt of public assistance than were legally involved girls ($\chi^2_{1} [N=114]=7.39, P=.007$).

Most girls (74/114, 64.9%) provided at least one social media account as a study contact method (range 0-3; Figure 1); 70% (52/74) provided 1, 27% (20/74) provided 2, and 3% (2/74) provided 3. Of those, 95% (70/74) provided account information from Instagram, 27% (20/74) from Facebook, and 11% (8/74) from Twitter. Results from individual chi-square tests showed that girls were significantly more likely to provide an Instagram versus Twitter ($\chi^2_{1} [N=114]=5.41, P=.02$) account, with a similar pattern for Instagram versus Facebook ($\chi^2_{1} [N=114]=3.54, P=.06$). Girls’ provision of at least one social media account did not differ by race/ethnicity ($\chi^2_{4} [N=114]=1.398, P=.85$), socioeconomic status ($\chi^2_{1} [N=114]=0.000, P=.99$), or age ($t_{112}=-0.608, P=.54$). Legally involved girls were significantly less likely to provide social media account information than school-referred girls (44% [16/36] versus 74% [58/78]; $\chi^2_{1} [N=114]=9.68, P=.002$).

Table 1. Demographic characteristics of adolescent participants (N=114).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adolescents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at baseline (years), mean (SD)</td>
<td>15.7 (1.3)</td>
</tr>
<tr>
<td>Race/ethnicity, n (%)</td>
<td></td>
</tr>
<tr>
<td>White, non-Hispanic/Latinx</td>
<td>14 (12.3)</td>
</tr>
<tr>
<td>Black, African American or Haitian, non-Hispanic/Latinx</td>
<td>22 (19.3)</td>
</tr>
<tr>
<td>Mixed race or multiracial, non-Hispanic/Latinx</td>
<td>18 (15.8)</td>
</tr>
<tr>
<td>Other, non-Hispanic/Latinx(^a)</td>
<td>15 (13.2)</td>
</tr>
<tr>
<td>Hispanic or Latinx</td>
<td>45 (39.5)</td>
</tr>
<tr>
<td>Legal system involvement, n (%)</td>
<td></td>
</tr>
<tr>
<td>Legally involved</td>
<td>36 (31.6)</td>
</tr>
<tr>
<td>School referred</td>
<td>78 (68.4)</td>
</tr>
<tr>
<td>Self or family receiving public assistance, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>37 (32.5)</td>
</tr>
<tr>
<td>No</td>
<td>77 (67.5)</td>
</tr>
</tbody>
</table>

\(^a\)Includes participants who self-identified as Asian, Native Hawaiian or other Pacific Islander, American Indian, or “Other.”

Figure 1. Social media account information provided by adolescent participants.
Discussion

The majority of school-referred and under half of legally involved adolescent girls (74% [58/78] and 44% [16/36], respectively) provided at least one social media account to researchers, suggesting that collecting social media information is possible among girls enrolled in substance use intervention trials. In comparison, national data on social media account ownership indicates that 72% of adolescents use Instagram and 51% use Facebook [19]. Social media ownership and utilization have not yet been examined among legally involved adolescents, yet adolescents belonging to minoritized racial and ethnic groups and adolescents with reported lower socioeconomic statuses, who are disproportionately represented in the juvenile legal system, demonstrate comparatively high social media usage [12,19,20]. Further, adolescents who use substances are likely to use social media [21].

Consistent with our hypothesis, legally involved girls were significantly less likely than school-referred girls to provide social media information. Legally involved girls’ lesser willingness may be associated with disruptions in phone or device access (eg, due to caregivers or court taking cellphones away; lack of Wi-Fi) or concerns about privacy (eg, fears of court, probation or diversion staff, or parents monitoring) and system mistrust (eg, posting potentially self-incriminating images, researchers not keeping information from the court). Future research should examine how these and other factors may impact adolescents’ willingness to provide social media information, especially among legally involved girls. Differences in willingness to share social media information and limitations to using social media for research purposes with adolescents who may be most vulnerable to legal system monitoring should also be examined.

Our findings also suggest that girls might have preferred platforms for contact (ie, Instagram). This is in line with US adolescent trends from 2018 that Instagram use has surpassed Facebook in popularity [19]. It will be important in future research to assess whether the higher frequency of platforms provided (eg, Instagram in this study) is due to more prevalent account ownership or to greater willingness to release Instagram information to researchers relative to other platforms. To account for emergent social media trends potentially unknown to researchers (eg, TikTok), researchers should allow participants to provide open-ended responses when asked about social media use and contact preferences.

The use of an existing data set not designed for the purpose of these types of analyses comes with limitations. For example, this data set did not include reasons regarding willingness to provide social media information. Data were also limited to adolescent girls, which is informative for much needed gender-responsive substance use intervention trials, but our data do not address how to reach and engage legally involved adolescents who do not identify as female in substance use intervention research, nor the impact of gender identity on willingness to provide social media information. Girls also provided social media information in the presence of a consenting caregiver, which may have impacted their choices to share social media information. Future research should explore adolescents’ perspectives on the use of social media for study retention purposes, assess feasibility and acceptability using standardized measures, and qualitatively examine factors impacting decisions to share accounts (eg, privacy from parents, courts). Ethical implications of collecting social media information from adolescents for substance use intervention research retention purposes should also be considered in future research.

Our findings provide some of the first empirical data demonstrating that social media information can be collected to reach and retain historically minoritized, underrepresented, and vulnerable adolescent populations (ie, girls, legally involved, racial/ethnic minority) in substance use intervention research. Understanding more about willingness to provide account information and patterns of use, and assessing the effectiveness of social media use on improving study outreach, recruitment, and retention are critical areas for future research.

Acknowledgments

The authors wish to thank their community, school and juvenile legal system partners for their collaboration, as well as the youth and families who agreed to participate in the research. This work was supported by the National Institute on Drug Abuse (NIDA) under grant numbers R01DA035231 (principal investigator: MT-S) and K24DA046569 (principal investigator: MT-S). ED’s effort was supported by R34DA050480. The NIDA did not have any role in the study design; collection, analysis, and interpretation of data; writing of the report; and in the decision to submit the report for publication.

Conflicts of Interest

None declared.

References


Collecting Social Media Information in a Substance Use Intervention Trial With Adolescent Girls With Lifetime Substance Use History: Observational Study

JMIR Form Res 2021;5(9):e25405
URL: https://formative.jmir.org/2021/9/e25405
doi:10.2196/25405
PMID:34505833

Lili M C Ramos, Joseline Delgadillo, Sarah Vélez, Emily Dauria, Jamie Salas, Marina Tolou-Shams. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.09.2021. 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.
A Suite of Mobile Conversational Agents for Daily Stress Management (Popbots): Mixed Methods Exploratory Study

Matthew Louis Mauriello, PhD; Nantanick Tantivasadakarn, BSc; Marco Antonio Mora-Mendoza, BSc; Emmanuel Thierry Lincoln, MSc; Grace Hon, BA; Parsa Nowruzii, BA; Dorien Simon, MSc; Luke Hansen, PhD; Nathaniel H Goenawan, BSc; Joshua Kim, BSc; Nikhil Gowda, MBA; Dan Jurafsky, PhD; Pablo Enrique Paredes, MBA, PhD

Department of Computer and Information Sciences, University of Delaware, Newark, DE, United States
Symbolic Systems Program, School of Humanities and Sciences, Stanford University, Stanford, CA, United States
Computer Science Department, College of Engineering, Stanford University, Stanford, CA, United States
Department of Information Systems, Institut Supérieur d'électronique de Paris, Paris, France
Stanford School of Medicine, Stanford University, Stanford, CA, United States
Alliance Innovation Lab, Silicon Valley, CA, United States
Department of Linguistics, School of Humanities and Sciences, Stanford University, Stanford, CA, United States
*these authors contributed equally

Corresponding Author:
Pablo Enrique Paredes, MBA, PhD
Stanford School of Medicine
Stanford University
291 Campus Drive
Stanford, CA, 94305
United States
Phone: 1 650 723 2300
Fax: 1 650 725 8381
Email: pparedes@stanford.edu

Abstract

Background: Approximately 60%-80% of the primary care visits have a psychological stress component, but only 3% of patients receive stress management advice during these visits. Given recent advances in natural language processing, there is renewed interest in mental health chatbots. Conversational agents that can understand a user’s problems and deliver advice that mitigates the effects of daily stress could be an effective public health tool. However, such systems are complex to build and costly to develop.

Objective: To address these challenges, our aim is to develop and evaluate a fully automated mobile suite of shallow chatbots—we call them Popbots—that may serve as a new species of chatbots and further complement human assistance in an ecosystem of stress management support.

Methods: After conducting an exploratory Wizard of Oz study (N=14) to evaluate the feasibility of a suite of multiple chatbots, we conducted a web-based study (N=47) to evaluate the implementation of our prototype. Each participant was randomly assigned to a different chatbot designed on the basis of a proven cognitive or behavioral intervention method. To measure the effectiveness of the chatbots, the participants’ stress levels were determined using self-reported psychometric evaluations (eg, web-based daily surveys and Patient Health Questionnaire-4). The participants in these studies were recruited through email and enrolled on the web, and some of them participated in follow-up interviews that were conducted in person or on the web (as necessary).

Results: Of the 47 participants, 31 (66%) completed the main study. The findings suggest that the users viewed the conversations with our chatbots as helpful or at least neutral and came away with increasingly positive sentiment toward the use of chatbots for proactive stress management. Moreover, those users who used the system more often (ie, they had more than or equal to the median number of conversations) noted a decrease in depression symptoms compared with those who used the system less often based on a Wilcoxon signed-rank test (W=91.50; Z=−2.54; P=.01; r=0.47). The follow-up interviews with a subset of the participants indicated that half of the common daily stressors could be discussed with chatbots, potentially reducing the burden on human coping resources.
Conclusions: Our work suggests that suites of shallow chatbots may offer benefits for both users and designers. As a result, this study’s contributions include the design and evaluation of a novel suite of shallow chatbots for daily stress management, a summary of benefits and challenges associated with random delivery of multiple conversational interventions, and design guidelines and directions for future research into similar systems, including authoring chatbot systems and artificial intelligence–enabled recommendation algorithms.

(JMIR Form Res 2021;5(9):e25294) doi:10.2196/25294

KEYWORDS

conversational agents; virtual agent; chatbot; therapy; stress relief; stress management; mental health; stress; exploratory; support; mobile phone

Introduction

Overview

In the United States, approximately 60%-80% of the primary care visits have a psychological stress component [1], but only 3% of patients receive stress management advice during these visits [2]. The reason for this is a combination of both limited infrastructure geared toward preventive health and limited focus on stress management. However, the increasing accessibility of mobile computing has spurred the growth of mental health apps, which currently account for 29% of the mobile health app market that includes fitness, nutrition, and other lifestyle apps [3]. However, general trends suggest that users are spending an increasing amount of time accessing services through messaging clients compared with purpose-built apps [3]. As a result, developers are leveraging these clients to build conversational interfaces, also known as chatbots, to create novel interactions in the health domain, including those that allow users to report symptoms, make appointments, and gain referrals.

Advances in natural language processing, such as intent [4] or emotional recognition [5,6] based on very large language data sets, continue to increase the range of these systems and their potential for impact. Research into improving conversational systems spans a number of domains such as customer service [7,8], companionship [9,10], and, increasingly, mental health [11-14]. As chatbots are scalable and easy to access, many systems are aimed at substituting human support in common conversations with known formats. Early efforts in mental health include ELIZA [15], which attempted to model the psychoanalytical approach of introspection: asking questions to engage the user in examining their own mental and emotional processes. More recently, chatbots such as Woebot [16] and Wysa [17] have been used to provide cognitive behavioral therapy (CBT) support to people at risk for depression. As a result, it is no surprise that a recent workplace survey found that most people (86% of those surveyed) were receptive to using chatbots and artificial intelligence (AI) systems that provide mental health support services [18]. However, given the complexity of life and the many types of stressors that a chatbot would need to understand to provide support, building a proactive everyday stress management chatbot is complex to design, costly to develop, and difficult to author in ways that appeal broadly.

To address these limitations, we aim to explore creating a new breed of simple conversational chatbots that use short conversations for in-the-moment management of daily stressors (eg, deadlines, difficult social interactions, and lack of sleep). Inspired by Etzioni’s second law for AI systems, “Disclose that it is not human” [19], we aim to create shallow yet effective and engaging mental health chatbots that do not try to replicate human intelligence. In the context of daily stress management, we define shallow chatbots as those that use few and brief conversational exchanges to deliver a single coping technique. These shallow chatbots are not created to replicate or replace humans (ie, family, friends, or therapists) but rather to operate as part of a larger ecosystem of agents providing stress management support. The advantages of creating multiple shallow chatbots are manifold: (1) chatbots capable of delivering microinterventions lower barriers of time and commitment for users; (2) they can be authored and curated more quickly by novice designers to produce a variety of high-quality advice options; (3) this variety of chatbots could help improve long-term engagement (ie, chatbots that fail could be removed); and (4) the suite approach allows for future personalization.

Prior research has explored the design of suites of just-in-time stress management interventions. For instance, the study by Paredes et al [20] demonstrated that a suite of microinterventions coupled with a web-based learning recommendation system could teach long-term stress coping skills to users. We extend this research on microinterventions by exploring a suite of diverse and specialized shallow chatbots for daily stress that we call Popbots. As early work investigating suites of shallow chatbots, our research questions are exploratory and include the following: How might we design multiple shallow chatbots for proactive and reactive stress management? How might everyday users react to using these multiple chatbots for managing their daily stress? And what challenges and benefits do they perceive about such systems?

Background

Daily Stress

The stress response is an evolutionary mechanism that mobilizes bodily resources to help humans cope with daily challenges as well as life-threatening situations. Stress has two components: a stressor and a stress response. The former could be linked to sources of uncertainty, complexity, cognitive loads, or emotional distress. The latter refers to the mental and physical reaction to such stimuli. Daily stressors are defined as the routine challenges of day-to-day living. The challenges can either be predictable (eg, daily commutes) or unpredictable (eg, an unexpected work deadline) and occur on 40% of all days. Unlike chronic stress, these stressors are relatively short-lived and do not persist from

https://formative.jmir.org/2021/9/e25294

JMIR Form Res 2021 | vol. 5 | iss. 9 | e25294 | p.78

(page number not for citation purposes)
day to day [21,22]. However, daily stress has been shown to exacerbate symptoms of existing physical health conditions [21]. Repeated triggering of daily stress can also lead to chronic stress, which has been associated with a variety of pathophysiological risks such as cardiovascular diseases and immune deficiencies—conditions that impair the quality of life and shorten life expectancy [23,24]. Thus, having effective mitigation strategies for daily stress can have a positive effect on a person’s well-being and overall health.

**Traditional Stress-Mitigating Interventions**

There is a wide variety of methods employed to help reduce stress. Positive psychology, for instance, is an emerging practice to help people calm down with personally targeted cues such as asking people to express gratitude or perform compassionate acts [25]. Another group of effective techniques is part of CBT [26], which teaches people how to recognize their sources of stress, change their negative behavioral reactions, and reframe their thoughts. Yet another approach is the use of narrative therapy, which focuses on constructing conversations to help people become satisfied with their state of being [27]. Such conversations are the basis of social interaction, which has a direct impact on emotions [28,29]. For example, positive social interactions have been shown to lead to calmness and openness in social engagement [29,30]. In our work, we borrow from this literature (ie, positive psychology, CBT, and somatic regulation) to design chatbots to guide users through stress-relieving techniques in response to daily stressors.

**Stress-Mitigating Microinterventions**

A relevant approach to this work is the use of internet-based technology that leverages specific aspects of CBT (eg, for smoking cessation [31,32]), positive psychology (eg, for depression [33,34]), and similar techniques to deliver personalized treatments and enhance well-being [35]. Recently, researchers explored the use of machine learning algorithms to recommend calming interactions with web apps. For instance, the study by Paredes et al [20] demonstrated the benefit of using just-in-time web-based interventions for teaching long-term stress-coping skills. In particular, the study discussed the complexity of engaging people to prevent early attrition. People under high levels of stress find that any additional task, including interventions, adds to their stress load. This motivates the need for research on the design of intervention suites that could reduce attrition by diversifying the types of interventions that are recommended to users over time [20,36,37].

**Chatbots for Mental Health**

Chatbots have a long history of application in mental health. The earliest mental health chatbot, ELIZA [15], was programmed to deliver nondirective therapy mirroring Rogerian therapy (ie, reflecting and rephrasing user input). A few years later, PARRY [38] was used to study schizophrenia. In addition to its capability of displaying regular expressions, PARRY included a model of its own mental and affect states. For example, PARRY could become more angry or mistrustful, thus generating hostile outputs. In a comparison study, psychiatrists could not distinguish transcripts of interviews with PARRY from those of interviews with people with schizophrenia. However, work on subsequent mental health chatbots did not emerge until recently [11-14].

Recent examples close to our work are varied and include chatbots that administer motivational stress management surveys [39] and CBT chatbots such as Woebot [16], Wyasa [17], and Tess [40]. Woebot is an automated chatbot based on the principles of CBT. Woebot leads users through a series of CBT-type lessons, directing users to videos and other forms of didactic material to get them to engage in common CBT skills such as cognitive restructuring or behavioral activation. Wyasa is an AI-driven pocket penguin that also bases chat interactions on CBT skills. The benefits of Woebot have been demonstrated in a randomized controlled trial showing superiority to a web-based e-book at reducing symptoms of depression and anxiety in a sample of college students, and a similar experiment was run with Tess, which corroborates these results across multiple university populations.

This expanding ecosystem of chatbots for mental health apps suggests that such tools are viable as accessible support solutions. This is not surprising, given that mental health has long relied on the talking cure as a primary form of treatment. A challenge regarding the use of existing chatbot systems is the need to explore the problems through a set of questions and answers and conversational exchanges that may be hard to author and maintain. Our system overcomes this limitation by allowing for the creation of multiple chatbots with each representing a single type of intervention. Authoring these shallow chatbots is easier for a designer because they can focus on delivering a single intervention technique with a clear objective and conclusion. For users, microintervention chatbots offer quick advice without their need to work through a lengthy dialog that could be, by itself, another source of stress. In some ways, our system resembles a game or movie platform (eg, Netflix) where each chatbot is a new game or movie and we can learn over time which chatbots the users prefer.

**Methods**

**Prototype Chatbot Suite**

Extending prior work on microinterventions and conversational interfaces [20], we propose the creation of a suite of shallow chatbots that provide in-the-moment conversations for managing daily stress. Although prior work tended to focus on patients or people at risk (ie, people with high levels of depression or anxiety symptoms as highlighted by recent surveys [11-14]), our aim is to provide a quick and engaging system using simple microintervention chatbots that can help to alleviate daily stress for healthy people (ie, toward improving long-term well-being and helping to mitigate future crises). Another goal of the project is to simplify the authoring of chatbots by reducing complexity toward enabling a scalable solution for rapidly creating numerous (ie, hundreds or more) chatbots for stress management. To explore this idea, we developed a prototype chatbot suite with a common template for short conversations (ie, 2-3 minutes with a few conversational exchanges) composed of four components (Figure 1): (1) an onboarding script for explaining the system and its limitations to users; (2) a shared code logic for handling user inputs; (3) a declarative model of the chatbot’s mental state, behavior, and response; and (4) a set of conversation rules that determine how the chatbot should react to different types of input.
set of greetings, stressor parsers, and intent-extraction components; (3) the microintervention chatbots that make up

**Figure 1.** Overview of conversation structure for all chatbots: When a user sends a greeting message (eg, “Hi”), they receive a greeting from the suite of chatbots in response. If it is the user’s first time using the suite, they are directed to the onboarding script explaining how our shallow chatbot suite operates and what its limitations are. Next in the conversation sequence, the system asks the user what it is that is currently “stressing them out”; the stressor is then extracted, and a chatbot is randomly selected. Each chatbot delivers a coping technique in the form of a brief conversation that ends with the user assessing the conversation on a 3-point Likert scale (ie, Not helpful, Neutral, and Helpful).

**Chatbot Design**

**Overview**

We used an iterative, human-centered approach to designing our chatbot suite (Table 1). The initial chatbot scripts were developed in a 4-hour workshop with the aid of 6 novice designers, curated by a clinical psychologist, and tested for quality purposes by conducting simulations in which pairs of designers acted as users and chatbots. Each chatbot relied on a decision tree to facilitate conversations, usually resulting in the user providing a response to a series of open-ended (eg, *What is the worst-case scenario for [a stressor]?")", yes-no (eg, *Has [the stressor] affected your sleep?")", or numerical (eg, *What is the severity of a scenario?") questions (Textbox 1). Stress management literature—particularly literature related to CBT techniques [26,41,42]—was used to derive conversations for stress relief. Using this approach, our novice design team created chatbots based on three techniques (ie, worst-case scenario, problem solving, and positive thinking). The total development time (ie, including design, curation, and quality assurance steps) was approximately 8 hours. We then evaluated the feasibility of our chatbot system against a control condition in a Wizard of Oz (WOZ) pilot study with 14 users (Multimedia Appendix 1). We observed that the participants in the condition with multiple chatbots tended to agree to a greater degree that the intervention helped to reduce their stress compared with those in the control condition with a single chatbot; however, follow-up interviews revealed that the participants still expected chatbots to act in human-like ways. The lessons learned from this pilot study were used to refine our chatbot scripts, and they also informed the development and implementation of our web-based system.

**Table 1.** Prototype chatbot names, their techniques, and the studies in which they were used.

<table>
<thead>
<tr>
<th>Chatbot</th>
<th>Technique</th>
<th>Description</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doom Bot</td>
<td>Worst-case scenario</td>
<td>Asks the user to consider the worst-case scenario</td>
<td>Wizard of Oz and web-based</td>
</tr>
<tr>
<td>Sherlock Bot</td>
<td>Problem solving</td>
<td>Asks a series of questions to pinpoint the problem</td>
<td>Wizard of Oz and web-based</td>
</tr>
<tr>
<td>Glass-Half-Full Bot</td>
<td>Positive thinking</td>
<td>Asks the user to view their problems in a new light</td>
<td>Wizard of Oz and web-based</td>
</tr>
<tr>
<td>Sir Laughs-a-Bot</td>
<td>Humor</td>
<td>Finds humor in the situation</td>
<td>Web-based</td>
</tr>
<tr>
<td>Treat Yourself Bot</td>
<td>Self-love</td>
<td>Reminds the user that it is all right to treat themselves</td>
<td>Web-based</td>
</tr>
<tr>
<td>Dunno Bot</td>
<td>Distraction</td>
<td>Asks user to think about events they are looking forward to</td>
<td>Web-based</td>
</tr>
<tr>
<td>Checkin Bot</td>
<td>Checking in</td>
<td>Asks whether the stressor affected daily activities</td>
<td>Web-based</td>
</tr>
</tbody>
</table>
Textbox 1. A sample chatbot script.

The script used by Doom Bot

- Tell me more details about [problem]?
- I’m sorry to hear that. What are you most afraid might happen as a result?
- Alright, on a scale of 1 to 10, 1 being impossible, 10 being certain, how likely is this scenario?
- Alright, in the case that this happens, what could you do to get back on track?
- Cool, looks like you have a plan B. Just remember, though you cannot control everything, there is a way to get back on your feet.

System Implementation

We implemented our chatbot suite in Telegram (Telegram Messenger Inc) [43], a data-security–compliant messaging platform, using a Python (Python Software Foundation) backend and a MongoDB (MongoDB Inc) database (Figure 2). Using prior experience and the observations obtained during the initial chatbot workshop, the research team generated 4 additional chatbots bringing the total to 7 and programmed the conversational scripts in Python. Interactions with these chatbots are automatic, rely on open text (as opposed to buttons), and are rule-based, using regular expressions to control the flow of conversations. Following our template, when the user messages the chatbots (ie, by typing “Hi”), they receive a friendly greeting message and are asked to describe their current stressor (Figure 2). After extracting the stressor, a chatbot is randomly recommended, and its avatar image is displayed (Figure 2). User input is passed to a state handler through the Telegram application programming interface; the state handler analyzes these data to generate a response. Once the response is generated, it is sent to the user, and the interaction is logged. After the conversation ends, the chatbot thanks the user for sharing and asks them for feedback on whether the interaction helped to reduce their stress on a 3-point Likert scale (ie, Not helpful, Neutral, and Helpful). We refined the chatbots with pilot users to make them seem more human-like (eg, introducing typing delays), clarified utterances so that users were more aware of when the system was waiting for input, and added a /switch option that allows users to change chatbots in situ (the only interaction that used buttons). Sample conversations are included in Multimedia Appendix 1.

Figure 2. A system diagram overview and example conversation scripts with conversational interfaces: (A) System diagram; (B) User who initiated a conversation over the Telegram interface being asked to describe their stressor; (C) Sample conversation with Doom Bot, recommended by the system.

Protocol

The participants were recruited in August-September 2019 through word of mouth and a university listserv. Our recruitment materials specified that participants would be asked to use our system for 7 days and complete a prestudy questionnaire, short daily surveys, and a poststudy questionnaire. These materials also specified that participants must be aged 18 years or older and have a compatible smartphone (ie, an Android [Google LLC] phone or an iPhone [Apple Inc]). Web-based enrollment occurred on a rolling basis, and all questionnaires were completed through the Qualtrics survey tool (Qualtrics LLC). After receiving our invitation email, the participants completed our prestudy questionnaire that asked them about their demographic information, how much stress they felt daily, and their perceptions of using chatbots for daily stress management. The participants also completed the short Patient Health...
Questionnaire (PHQ)-4 to ascertain a measure of their clinical anxiety and depression symptoms [44]. Upon completing the survey, the participants were automatically sent email instructions for installing the Telegram app as well as a personalized URL that, when accessed on their smartphones, initialized the Popbots channel within the app.

Once this initialization was completed, the participants were instructed to type “Hi” and go through the onboarding script that explained the purpose of the system (eg, that it was for daily stress management) and its limitations (eg, that it was not intended for the treatment of serious mental health conditions). After going through the onboarding script, the participants were instructed to interact with the chatbots anytime they felt stressed over the next 7 days. Daily surveys, which were sent at 8 PM each day (local time), asked the participants to rate their daily stress levels, sleep quality the previous night, and level of social interaction experienced that day. After 7 days of using the system, the participants completed a poststudy questionnaire, which asked the participants about their perceptions of daily stress over the course of the week and if their perceptions of chatbots had changed, as well as other usability questions. The participants also completed the PHQ-4 questionnaire again. We then followed up with a subset of the participants to complete a semistructured interview and card-sorting task (similar to our pilot WOZ study); we sent a general email request to all participants, and volunteers were enrolled on a first-come, first-served basis.

To motivate participation, we provided compensation. The participants earned US $10 through an Amazon gift card (Amazon Inc) for successfully completing both the pre- and poststudy questionnaires. We offered an additional US $3 for each day that they interacted with the chatbots and completed the daily survey. Compensation was prorated based on partial completion of these components. The participants who were interviewed after the study were compensated with an additional US $25 per hour of the interview. The protocols were reviewed for ethics and privacy concerns by our institution’s research compliance office.

**Participants**

We recruited 47 participants (34 women and 13 men). Most (33/47, 70%) were university staff members, whereas the remaining participants (n=14) were undergraduate students. Although the staff members were aged between 18 and 74 years, the students were aged between 18 and 24 years (Table 2). Approximately half of the participants (21/47, 45%) identified themselves as Asian, whereas the remaining participants identified as White (n=12), Hispanic or Latino (n=7), Multiracial (n=2), Black (n=2), American Indian (n=1), or preferred not to identify their race (n=2). More than half (28/47, 60%) reported being single (with no children), less than half (n=18) were married or in a domestic partnership (mean 1.9 children), and 1 participant was separated (with 3 children). Although the students had completed high school or General Educational Development requirements and were now working on their bachelor’s degree, the staff members had a high degree of formal education, with more than a third holding a bachelor’s degree (13/33, 39%), just less than a third (n=9) holding a master’s degree, some (n=4) having some college course experience, and a few more (n=4) holding terminal professional or doctoral degrees. Most of the staff members (30/34, 88%) were employed full time, whereas the remaining (n=4) were working part time; all student participants (14/47, 30%) listed their occupation as full-time students. Excluding interview payments, the participants received an average of US $23.86 (SD US $14.35; median US $25.00) in compensation.

### Table 2. Participant age ranges by subpopulation.

<table>
<thead>
<tr>
<th>Population</th>
<th>18-24 years, n (%)</th>
<th>25-34 years, n (%)</th>
<th>35-44 years, n (%)</th>
<th>45-54 years, n (%)</th>
<th>55-64 years, n (%)</th>
<th>65-74 years, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students (n=14)</td>
<td>14 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Staff members (n=33)</td>
<td>5 (15)</td>
<td>6 (19)</td>
<td>9 (26)</td>
<td>6 (19)</td>
<td>5 (15)</td>
<td>2 (7)</td>
</tr>
</tbody>
</table>

**Data and Analysis**

In summary, our data include responses to pre-, daily, and poststudy questionnaires; application logs from the chatbot system; interview transcripts; and photographs of the system; interview transcripts; and photographs of the chatbot system. We used the letter P and randomized IDs to refer to the participants in our web-based study (eg, P1234) and letters (eg, PX) to refer to those from our WOZ pilot study in our interview results.

https://formative.jmir.org/2021/9/e25294
Results

Application Logs
Over the course of 7 days, most of the participants (44/47, 94%) interacted with our chatbots, generating 291 conversations. The participants averaged approximately seven conversations per week (mean 6.83, SD 3.14). These conversations were short, lasting only a few minutes (mean 1.95, SD 2.53), and often occurred during the latter part of the day. Although some conversations were likely triggered by the daily survey reminder (at 8 PM), most (232/291, 79.7%) of the conversations were unprompted and occurred throughout the day with increased activity in the 7 AM, 12 PM, 3 PM, and 8 PM hours. A deeper exploration of these conversations indicated that some participants were simply checking in, particularly around 8 PM, reporting stressors such as “Nothing” or “Doing pretty good actually.” As a result, we filtered out approximately a third of the conversations that fell into this category as well as those that contained a technical issue making them indecipherable.

Reporting Stressors
We observed 2 ways that the participants reported stressors to the chatbots. Most of the participants (146/197, 74.1% of conversations) tended to describe stressors in a few words. For example, participants wrote “Having to go to work tomorrow,” “My presentation that’s coming up,” and “My friend being mad at me.” Another approach (51/197, 25.8% of conversations) was to type out single words (eg, “Money,” “Car,” and “Family”).

Topics of Conversation
After filtering out erroneous and nonstress-related conversations, we labeled the remaining 197 conversations using eight category tags representing the consistent topics that the participants discussed with the chatbots (Table 3). The most common topics included (1) work- and school-related productivity issues, (2) health problems (eg, feeling tired and experiencing pain), and (3) interpersonal issues related to (nonfamilial) social relationships. There were also a number of Other conversations that were not widely discussed but might point to additional topics of daily stress, including vacation-related stress (eg, packing), commuting, and seasonal stressors (eg, holiday-related gift giving).

Table 3. Categories of stressors.

<table>
<thead>
<tr>
<th>Stressors</th>
<th>Examples</th>
<th>Count (n=197), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work, school, and productivity</td>
<td>“I have some tasks I keep putting off”</td>
<td>79 (40.1)</td>
</tr>
<tr>
<td>Health, fatigue, and physical pain</td>
<td>“I want to eat better but I’m having a hard time with it”</td>
<td>27 (13.7)</td>
</tr>
<tr>
<td>Social relationships</td>
<td>“I found out my ex has a new girlfriend”</td>
<td>21 (10.6)</td>
</tr>
<tr>
<td>Financial problems</td>
<td>“I have a friend coming by and I’m stressed about an expense”</td>
<td>13 (6.6)</td>
</tr>
<tr>
<td>Emotional turmoil</td>
<td>“Feeling lonely”</td>
<td>12 (6.1)</td>
</tr>
<tr>
<td>Family issues</td>
<td>“My marriage”</td>
<td>10 (5.1)</td>
</tr>
<tr>
<td>Everyday decision-making</td>
<td>“Don’t know what to cook for dinner”</td>
<td>8 (4.1)</td>
</tr>
<tr>
<td>Other</td>
<td>“Just travel stuff”</td>
<td>27 (13.7)</td>
</tr>
</tbody>
</table>

In Situ Efficacy
Overall, in situ efficacy was either helpful (76/197, 38.6%) or neutral (64/197, 32.5%). As it is reasonable to expect that not all interventions will be viewed as helpful, a neutral response may also be viewed as positive in terms of a system with multiple chatbots. However, more concerning is that the remaining conversations were rated as unhelpful (57/197, 28.9%). We also observed that feedback varied by chatbot (Figure S1 of Multimedia Appendix 1). For example, nearly half of the conversations of Treat Yourself Bot were rated as helpful and most (22/31, 71%) were rated positive or neutral versus those of Checkin Bot, which were mostly viewed as unhelpful (13/25, 52%). We believe that these results are encouraging because they suggest that with more data, patterns between stressors and chatbot or user and chatbots may emerge that might explain these differences and allow a future system with more complex recommendation algorithms to learn from and make personalized recommendations for each user-stressor pair.

Daily Surveys
The daily survey was administered each day at 8 PM (local time). In addition to usability questions, the survey tracked levels of stress, social interaction, and sleep quality the previous night using 5-point Likert scales ranging from None to Very high or Very poor to Very good. As a third (67/197, 34%) of the conversations could not be matched to a daily survey (ie, because the participants did not complete them that day), our analysis focused on evaluating trends in the matched conversations (n=130). With respect to general trends, we noted that most (91/130, 70% of conversations) were matched to surveys that reported Low to Moderate levels of stress throughout the week. Similarly, most (114/130, 87.6%) were matched to surveys reporting Acceptable or better sleep quality the night before and most (110/130, 84.6%) to surveys reporting Low to High levels of social interaction each day. However, conversational feedback was roughly evenly distributed across these variables, skewing slightly toward Helpful. Further analysis did not reveal any strong correlations between these variables and in situ feedback.
Poststudy Experiential Feedback

Open-ended feedback from the poststudy questionnaire was generally positive and helped to characterize the participant experience. For example, although it seemed from the application logs that the participants were using the chatbots throughout the day, most considered using the chatbots to be a private activity and, as a result, reported that they were difficult to use in the moment. Most of the participants (28/31, 90%) reported using the chatbots when they were alone—typically when they had a free moment (ie, a few hours after a stressful event). This was often because in their work and social environments, they were busy or wanted to avoid giving the perception of rudeness caused by being on their phones, which is an interesting potential barrier.

Like the in situ conversational feedback, retrospective feedback on effectiveness was skewed positive. Most (25/31, 81%) of the participants viewed the chatbots as Slightly effective to Very effective, and approximately one-fourth (9/31, 29%) described the chatbots as Not effective at all. Approximately half (17/31, 55%) of the participants described the current set of chatbots as cute and engaging. They also appreciated the concept of having a variety of chatbot options available. A participant explained as follows:

I like the ability to have access to different chatbots. I liked problem solving bot and Checkin-bot, but the laugh bot not so much. [P7596]

However, with only 7 chatbots available, some (n=6) of the participants commented that their interactions with the chatbots felt formulaic or repetitive.

Pre- and Poststudy Comparison

Post Hoc Analysis

As part of our analysis, we looked at changes in several questions asked across the pre- and poststudy questionnaires. These pre/post metrics include changes in the PHQ-4 scores, perceptions of daily stress, and perceptions of chatbots for stress management. To further explore these differences, we also conducted a post hoc analysis. We separated users into 2 groups based on the number of conversations that the participants had had with the chatbots. Specifically, we grouped participants whose completed number of conversations was less than or equal to the median number of conversations into the Low use group and the remaining into the High use group. The participants in the Low use group (n=16) had an average of 4.31 conversations over the course of the week (SD 1.31), whereas the participants in the High use group (n=15) had twice as many conversations (mean 8.67, SD 2.12).

PHQ-4 Scores

Overall, we observed a decrease in PHQ-4 scores over the course of the week when comparing pre- and poststudy assessments for the participants who completed the study (Figure S2 of Multimedia Appendix 1). The medians of the before-and-after PHQ-4 scores were 3.0 and 2.0, respectively. A Wilcoxon signed-rank test showed that this decrease was significant (W=91.50; Z=-2.54; P=.01; r=0.47). Although we cannot directly attribute this decrease to the interactions with our chatbots without a control group, our post hoc analysis suggests that although the scores of both groups for the prestudy PHQ-4 were similar (median 3.0), there was a greater reduction in the PHQ-4 score (median 2.0) of the High use group, which was significant (W=18.0; Z=-2.16; P=.03; r=0.57), compared with that of the Low use group (median 2.5), which was not. We theorize that these data point to the potential efficacy of our approach.

Daily Stress Experience

We evaluated perceptions of daily stress using a 4-point Likert scale ranging from A little to A great deal. Although the participants reported varying levels of stress on the daily survey, most described their perceptions of daily stress as Moderate in the prestudy questionnaire, and the perceptions of daily stress after their participation were retrospectively similar (Figure S3 of Multimedia Appendix 1). Although we observed a slight decrease in perceived daily stress, these changes were not significant.

Perceptions of Chatbots

When asked to describe their perceptions of chatbots for stress management on an open-response question, approximately half of the participants (22/47, 47%) were neutral (ie, they stated that they had no opinion on chatbots); slightly more than a third (n=17) were positive (ie, they believed that chatbots could be helpful); and the remaining (n=8) participants were negative (ie, they believed that chatbots would not be effective). An illustrative comment in favor of chatbots was by P8530: “They seem to be a viable option for the management of stress, but they need to be further refined in order to be useful in day-to-day situations.” In contrast, those who were more negative were best exemplified by P5219, who wrote: “…it doesn’t seem like talking to a non-human would be that helpful because, for me, talking to a human doesn’t usually help.”

However, in the poststudy questionnaire, most (20/31, 65%) of the participants reported a more positive attitude toward chatbots for mental health. This was often because (1) they had a positive experience with the system themselves, (2) they could conceive of such systems being helpful to people more generally, or (3) they found the activity of taking some time out each day to think about their stress helpful. In addition, approximately half (16/31, 52%) of the participants agreed that they had learned something about stress management from interacting with the system. For example, P8002 noted, “I liked the idea of congratulating yourself for the things you did manage to do rather than focusing only on what you didn’t.” Interestingly, even the participants who did not report learning anything from the system were positive. For example, P9907 noted that although they did not learn anything from the interactions with the chatbots, they were “helpful reminders of what I should be doing when I am stressed.” Others noted that although they did not learn anything directly from the chatbots, they did learn that chatbots could be effective tools. A third of the participants (n=10) reported no change in their general attitudes toward chatbots, and a small number (n=3) reported a more negative attitude (ie, they found the chatbots too repetitive or poorly implemented).
Follow-up Card-Sorting Interviews

Two Phases

The interviews primarily centered around a card-sorting activity with two phases. In the first phase, the participants (N=13) were given 13 stressors to be assigned to the different chatbots based on the chatbots that they felt were most effective. The stressors were synthesized from the Holmes and Rahe Stress Scale [46]. In the second phase of the activity, the participants were asked to redistribute the stressor categories among three additional human options in addition to the chatbots: a nontrained stranger, friends and family, and a therapist. The participants were asked to think aloud while making their assignments.

Card-Sorting Results

The card-sorting activity suggested that there were certain stressors that the participants preferred to talk to the chatbots about, given that not all stressor categories were reassigned in phase 2 when humans were available. We observed that 47% (79/169) of the stressors were retained by the chatbots (Figure S4 of Multimedia Appendix 1). This result, we believe, is critical and points toward a willingness by the participants to use the chatbots for common daily stressors.

Moreover, when we sort these stressors by those most assigned to the chatbots, we observe that Everyday decisions and Financial stress were rarely reassigned to humans, whereas interpersonal issues such as Romantic stress or Conflict with family and complex topics such as Sexuality and identity were. However, not all chatbots performed equally well in terms of retaining their assignments in the presence of humans. For example, Table 4 indicates that Checkin Bot, Sherlock Bot, and Doom Bot were some of the more resilient chatbots, whereas most of Dunno Bot’s assignments were reassigned to humans. In fact, many chatbots retained more than half of their assignments. We also noted that the participants had a strong preference for assigning problems to Friends and family over therapists, with two assignments made to strangers.

Table 4. Stressor assignments by chatbot and human resource (n=169).

<table>
<thead>
<tr>
<th>Resource: chatbots and humans</th>
<th>Stressor assignments, n (%)</th>
<th>Phase 1: chatbots</th>
<th>Phase 2: chatbots and humans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sherlock Bot</td>
<td>45 (26.6)</td>
<td>21 (12.4)&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Glass-Half-Full Bot</td>
<td>30 (17.7)</td>
<td>5 (2.9)</td>
<td></td>
</tr>
<tr>
<td>Doom Bot</td>
<td>23 (14)</td>
<td>12 (7.1)</td>
<td></td>
</tr>
<tr>
<td>Sir Laughs-a-Bot</td>
<td>21 (13.6)</td>
<td>12 (7.1)</td>
<td></td>
</tr>
<tr>
<td>Treat Yourself Bot</td>
<td>20 (11.8)</td>
<td>11 (6.5)</td>
<td></td>
</tr>
<tr>
<td>Dunno Bot</td>
<td>15 (8.9)</td>
<td>5 (2.9)</td>
<td></td>
</tr>
<tr>
<td>Checkin Bot</td>
<td>15 (8.9)</td>
<td>14 (8.3)</td>
<td></td>
</tr>
<tr>
<td>Friends and family</td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
<td>59 (34.9)</td>
<td></td>
</tr>
<tr>
<td>Therapist</td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
<td>29 (17.2)</td>
<td></td>
</tr>
<tr>
<td>Stranger</td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2 (1.2)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Values in italics indicate that the percentage of the total decreased compared with phase 1 when human resources were unavailable.

<sup>b</sup>N/A: not applicable.

Qualitative Insights

Important Themes

As the participants made their assignments of stressors to the available chatbots and human resources, we probed for their rationale. Overall, we corroborated important themes around the desire to have chatbots that are part of an ecosystem of support supplementing humans, behave in a human-like way, and are available to discuss certain stressors.

First Impressions

A challenge with chatbots is that of first impressions. Approximately half of the participants (6/13, 46%) thought that their first interaction with a chatbot had an impact on their overall perceptions of the multiple chatbots available, and an unpleasant first interaction with a chatbot left the participants with a negative impression. For example, I participant stated as follows:

I went on the app and the bot said, “Find a joke” and it was something actually really terrible that was going on. That was my first time interacting with the bots. I thought “Wow, there’s nothing that’s funny about this.” This is not helpful at all. [P1962]

Benefits of Multiple Chatbots

The participants described several benefits of having multiple chatbots available, including the ability to combine two or more chatbots to address a problem. This point was raised during our WOZ experiment by PB, who was insistent that problem solving is ultimately the solution to all stressors, although other interventions may be used before, or in conjunction with, problem solving for better results:

Everything is going to end up here in problem solving. If people are calm and collected, then they can think well. So, if people are calm first, then they will find everything’s fine. Sometimes you can go from extreme
stress to humor, but that’s a big jump. I think it’s better if you’re slightly calmer and then humor comes in and then distraction. [PB]

We probed this idea further during the latter phases of our web-based study. Nearly half (6/13, 46%) of the participants agreed that using chatbots (or interventions) in combination could be an effective strategy to address stressors. For example, several were interested in using other interventions in combination with problem solving:

In the case of conflict with a coworker, distracting yourself, not letting it take over your life, looking at the positive side of things could help. It could also go to the treat yourself. And then the worst-case scenario, ‘Sure, I no longer interact with this coworker, and that’s okay.’ In the end going back to the Problem Solving. [P5279]

However, 1 participant (P5981) noted that while interacting with more than 1 chatbot can be helpful to address a problem, it is not necessary to use them in rapid succession.

Talking With Friends and Family
Most of the participants (11/13, 85%) favored talking with friends and family over talking with chatbots in some cases, and they indicated that this preference had to do with the complexity of the stressor. The participants preferred speaking with friends and family about difficult emotional problems (eg, conflicts with coworkers or interpersonal relationships). A participant summarized as follows:

It depends on the degree of the problem. If it is a huge problem, I want a real person. If it’s medium to small problem, then I go to the bot. [P1442]

There were several reasons for this preference, including relationship history and range of responses. Friends and family already have pre-existing relationships with the participants and knowledge about their personal lives. Approximately a third (n=4) of the participants preferred humans because they can show empathy. Another third (n=4) believed that humans are better at problem solving.

Talking With Therapists
Similar to talking with friends and family, more than half of the participants (7/13, 54%) said that they believed that therapists would be more helpful than chatbots in resolving complex problems. A participant observed as follows:

Therapists are trained and objective. They are actual people. You can have complex conversations and get answers to questions with them. [PC]

For example, nearly half (3/7, 43%) of the participants believed that a therapist would be very helpful for talking through issues of sexual identity.

Talking With Chatbots
The participants noted several practical and emotional benefits of talking with chatbots. Regarding practical reasons, most (11/13, 85%) of the participants suggested that chatbots have some advantages over humans. Almost half (n=5) of the participants mentioned that talking to a chatbot could help them avoid putting an undue burden on others. A participant stated as follows:

[Work stress] can be in the middle of the day, and [my friends] are going to be busy, and I don’t want to text them and bother them about that [P7596]

Similarly, some (n=3) of the participants also noted that chatbots are easy to access:

It’s going to be a lot quicker to pull up an app, right? I sneak away to a room, I pull up the bot app, it’s a lot quicker than messaging someone like, ‘Hey, are you around?’ and then waiting for a message back, or calling someone. [P7596]

Another reason cited by a few (n=3) of the participants was that they could more easily control how much they told chatbots, whereas humans are more likely to press for information.

Regarding emotional coping, the participants explained that the chatbots allowed them to shift their thinking about their stressors. For example, more than half of the participants (8/13, 62%) reported that Doom Bot helped them to recalibrate the gravity of their stressor:

It’s nice to hear when it feels like you’re on the brink of doom, that like, oh, this is the worst thing that can happen. [P5279]

Other participants (n=4) mentioned Glass-Half-Full Bot as being effective for putting stressful events in a different light. A participant, PD, shared that reflecting on the positive aspects of their experience allowed them to “take the edge off and make [the situation] work.” Similarly, half (n=7) of the participants described the chatbots as a distraction from their problems, potentially because of the immersive nature of conversational interventions. Moreover, almost half of the participants observed that humor helps them ameliorate their stress; as 1 participant, P7616, stated, “Humor is often the antidote.” These participants noted that chatbots with amusing dialog could be especially effective for stress management, although humor is highly subjective and thus difficult to make sure it appeals to everyone.

Privacy and Trust
When the participants were asked about any privacy concerns that they had about the platform, they were split. Approximately half (6/13, 46%) of them found some topics too personal to discuss with friends and family, but they were open to talking to chatbots because of the perceived privacy they provide. For example, a participant noted as follows:

I’m a very private person. I don’t like to talk about a lot of things even with friends and family or in therapy. [P1962]

Others went so far as to say that chatbots were more trustworthy because, as P7596 stated, they are “devoid of things that come with being human-like judgment or telling secrets.” In contrast, a few (n=4) of the participants noted that they were aware that their messages were not private and took comfort in knowing that therapists were ethically bound to keep conversations confidential. The remaining participants (n=3) were unsure:
I don’t know whether to worry about privacy or not. I think I have brand loyalty, so I always feel like Apple is gonna keep my stuff private. [PD]

When time allowed, we probed a bit more on this topic to get a sense of how users felt about chatbot systems using their data to improve the systems’ usefulness, and 2 concerns emerged. First, approximately a third (4/13, 31%) of the participants expressed concern about the use of conversational logs and other metadata that can be collected about web-based experiences. For example, P1962 likened such systems to other technology-related privacy incidents, stating, “even though I found the chatbots helpful, if they were like [Amazon’s] Alexa, running in the background waiting and listening to you and recording everything, I wouldn’t like that.” One-fourth (n=3) of the participants were concerned that, even with additional training, chatbots might not be able to be trusted to handle mental health crises (eg, referring users to proper resources). As P6716 summarized, “Chatbots should potentially set off an alarm and say there needs to be a human to prevent this person from doing something terrible, as opposed to just being an ultra-safe communication cocoon.” In contrast, 2 participants were unconcerned about the handling of their data as long as it was used to improve their experience. As P5219 stated, “I’m okay with chatbots having a lot of data about me if it’s going to help them to respond better.”

Discussion

Principal Findings

In this work, we explored the potential effectiveness and user perceptions of a suite of multiple chatbots for the management of daily stress in a web-based study. Our results suggest that multiple shallow chatbots, grounded in CBT and other techniques, can be designed quickly by relatively novice designers and that these chatbots could have a positive impact on mental health and well-being. We draw these conclusions from the observations that the in situ feedback indicated that most conversations were viewed as helpful, or at least neutral, and that there was a reduction in the PHQ-4 scores. As a complement to these results, there was the general positive improvement in sentiment toward the effectiveness of chatbots for daily stress management as well as other qualitative feedback that was consistent with these conclusions. However, participant bias is a concern when evaluating this feedback.

Although this study lacked a direct control, the fact that there seemed to be a bigger reduction in the PHQ-4 scores in those who used the system more often is encouraging, given that the users also perceived the level of daily stress that they experienced during participation as consistent with their prestudy perceptions of daily stress. Although we did not observe a reduction in these retrospective perceptions of daily stress, it is unlikely that such perceptions would shift, given the duration of the study and the general health of the population (ie, most of the participants reported sleeping well, being social, etc). Moreover, although many participants were positive about the variety of the chatbots in our suite, some indicated that it was not necessarily the conversations that they had had that were helpful but rather the act of taking the time out to reflect. Although we generally make no distinction between a chatbot in our system and a microintervention, the act of taking some time out is itself a microintervention, regardless of the user’s feedback on the chatbot that they were paired with. From the perspective of our system, either outcome is acceptable if users are engaged with the system and this engagement results in users being better equipped to manage daily stress. Next, we discuss some additional observations and opportunities enabled by our work, as well as its limitations. We close with design recommendations and discuss areas of future work.

Ecosystems of Support

When the participants were asked to assign the stressor categories to available human and chatbot resources in our follow-up card-sorting tasks, it was interesting to observe that nearly half of the daily stressors assigned to the chatbots remained with them when humans are also available. This suggests that, in the context of proactive stress management, the participants viewed chatbot systems such as ours as expanding the ecosystem of available support. This observation alone is critical because it suggests the potential of our system, and of chatbots more generally, to help with proactive stress management. The impact of a successful implementation could greatly increase access to stress management advice with a potential downstream impact of improving users’ well-being and helping to mitigate future crises. When the participants were asked to explain the rationale behind these assignments, they stated that they viewed chatbots as most effective for coping with low-complexity stressors (ie, practical and day-to-day concerns) compared with high-complexity stressors (ie, those of a more social or interpersonal nature) because of the relative ease of accessing chatbots and the perception of privacy granted by such systems. Another benefit that the participants perceived about their experience with our suite of chatbots was the potential to reduce the burden on available human-provided coping resources, which was also observed with other mental health chatbots used in long-term care [47].

Lowering Barriers to Authoring Chatbots

The participants who completed the study were positive about their use of the chatbot suite. Some benefited by learning new coping techniques (eg, positive reframing) and others by being reminded to take a moment out of their day to reflect. Although our results are preliminary, we believe that this interest in using a variety of chatbots for different problems or using multiple chatbots in sequence (as reported by some users) could improve engagement and help prevent attrition in chatbot systems for mental health—a problem observed in recent studies [48]. Although our implementation was also faced with these problems, a key difference in terms of solutions is that authoring new content in our suite means simply authoring another shallow chatbot, which can be done rapidly, whereas authoring new content for a single chatbot system must be done in a way that matches existing traits (eg, personality) and norms, which can be a limiting factor. The suite approach, we believe, sets an expectation of new and different content, decreasing this burden on design and offering more opportunities to appeal to different users. This unique solution could reduce the complexity and cost of developing chatbots for mental health by shifting focus.
to simpler chatbot designs. However, our work also suggests that relatively simple chatbots that disclose that they are not human should still appear human and converse in human-like ways (eg, show empathy [49]) if they are to be accepted and engaging.

If future controlled and longitudinal studies demonstrate that suites of shallow chatbots can be effective, then another long-term benefit of this approach could be the democratization of chatbot design, which is dominated by professionals who are highly trained in user experience, linguistics, and other fields. In contrast to the narrative that experts know best, in *Democratizing Innovation*, Von Hippel [50] argues that users generate significant design innovations more effectively than experts because they are highly motivated to solve their problems and share solutions. If we can design tools and methods that make authoring chatbots easier (eg, reducing the need to understand complex linguistic topics), then we can greatly reduce barriers to authoring effective chatbots in both multiple- and single-chatbot scenarios. We envision a future where anyone from everyday users to clinicians looking to augment or supplement their practice can author and recommend shallow chatbots to others as an immediate coping resource for daily stress. However, this raises the question of challenges that need to be addressed, including reducing the need to learn complex conversational design tools.

**Recommendation Systems**

In this study, chatbots were recommended to users at random, but an alternative approach could be the use of a recommendation system. Although conversational feedback skewed positive, some chatbots performed better than others (ie, feedback was more positive), and we theorize that installing a reinforcement learning algorithm that can better match a shallow chatbot to the user’s stressor could improve feedback further. For example, we noticed in some of our conversational logs that Doom Bot, which asks users to think about a future worst-case scenario, is not always appropriate for dealing with problems that exist in the past. Moreover, as our collection of shallow chatbots increases, it will be almost impossible for users to select an appropriate chatbot themselves, given the potential multitude of chatbots enabled by our less cost-intensive authoring paradigm. Similar to the study by Paredes et al [20], such algorithms can better take into account contextual, conversational, and prior interaction data to improve the matching between user problem and shallow chatbot, potentially personalizing to the user’s specific preferences over time. Chatbots that perform well across users could help solve the first-impressions challenge raised by our participants, which is an interesting solution not afforded to single chatbot systems [8]. In contrast, chatbots that generally perform poorly will not be recommended and thus could be discarded. In the future system we are working toward, users would be able to author many shallow chatbots quickly and deploy them, and the recommendation system would play an important dual role: recommending and curating appropriate and efficacious shallow chatbots that fit user context and stressor.

**Ethical and Privacy Considerations**

In our follow-up interviews, the participants raised concerns related to privacy and ethical responsibilities, which our system shares with other chatbot systems. As has been observed in recent studies, this includes preserving user privacy, detecting what problems the system can handle and when escalation to a human is necessary, and clearly describing the limitations of the system and ensuring that such systems are safe to use [51,52]. Although we did not necessarily discuss this topic with all participants, we learned that participants vary in terms of their understanding of, and preferences toward, privacy. Some trusted the system to remain private, whereas others knew that researchers and developers would use these data to make improvements. Still others pointed out that the chatbots allowed them to control the amount of information that they needed to divulge to mitigate their stress, which was appealing and certainly suggests that detecting something like escalating a problem could be quite challenging when less detailed information is being provided. This last point, we believe, is interesting because users do have an agency, which should be respected, but it is clear from other domains that explaining permissions and limitations of web-based systems is a challenging topic to get right [53]. As our system grows, we will increasingly need to accommodate differing privacy preferences and levels of agency with respect to important concerns such as user safety.

**Design Recommendations**

On the basis of this study, researchers and app designers engaged in designing multiple chatbots with a similar architecture might benefit from considering the following design recommendations:

1. Focus on lowering barriers to authorship and generating numerous shallow chatbots based on the vast amount of available psychological interventions for stress management.
2. Design for learning algorithms to handle recommendation and curation of interventions.
3. Attempt to score, rank, and classify daily stressors before assigning chatbots (interventions) to accommodate the differences in low- and high-complexity stressors as well as concerns about identifying problems that are too severe for the system to handle.
4. Consider a multitude of user coping and conversational styles, including users who may need a guided intervention or just an opportunity to reflect by talking or typing it out into the void.
5. Measure user personality, chatbot efficacy, and system engagement to optimize interactions across users.

If these problems can be addressed, then there is a real possibility of using this design paradigm to create a new breed of shallow chatbot systems that might be more engaging over the long term. However, the most difficult task is to convey the utility of these shallow chatbots to potential users for daily stress management. For the Popbots, our target group is healthy people regularly undergoing daily stress who are less likely to use preventive health systems. These users are a relatively understudied population in mental health, making research into engaging them and allowing them to explore the different
Limitations and Future Work

In addition to the aforementioned items, some additional limitations of this work include that the population in both studies was small and limited to students and staff members of a single university, which is likely not representative of the general population. Moreover, the population consisted largely of women, thus introducing a potential gender skew. The field studies were conducted during a single week, which is not sufficient to capture long-term effects, and, despite its privacy advantages over other platforms, Telegram is not a common messaging app. Downloading this app represents a significant barrier to adoption and may have contributed to attrition (eg, 4/44, 9% of participants registered for the study but did not sign in to Telegram). In addition, the compensation schema used to reduce attrition also incentivized the creation of off-topic data and likely influenced participant behavior. Future work should focus on monitoring and providing feedback about intrinsic improvements and avoid extrinsic incentives.

In the short term, we plan to improve the modularity of our suite design, explore the possibility of adding reminders within the system to improve consistency in use, and implement additional user-experience improvements, including the introduction of new chatbots that explore a larger range of interventions (eg, somatic breathing). To address the limitations of population and timescale in future evaluations, we aim to conduct a randomized controlled trial with a larger sample of diverse participants over a 4- to 8-week period with an appropriate control group and explore additional evaluation metrics that will make comparing the system with others easier (as suggested in the study by Abd-Alrazaq et al [54]). Using the data from this study, we plan to create a web-based learning recommendation system that helps pair users to our Popbots, given their stressor and context. An extension of this idea is creating an algorithm to detect whether the Popbots can handle a particular stressor and referring users to additional resources if needed (eg, calling 911 or seeking specialized help). To create an ecosystem of support with chatbots, we also plan to develop an authoring tool that will empower both mental health professionals and everyday users to create an increasing number and variety of chatbots, allowing us to compare their performance in numerous ways (eg, with other chatbots and by author type).

Conclusions

In this study, we have presented Popbots—a suite of shallow just-in-time chatbots that help users deal with daily stress. The system is scalable and provides variety in delivering numerous interventions rapidly, preventing attrition. We conducted multiple exploratory studies on the use of these microintervention chatbots for daily stress management, including a WOZ feasibility study, which we used to justify our approach of using multiple chatbots. We then iterated on the design of this system and tested its efficacy in a web-based pilot study. The results indicated that the users experienced a decrease in depression symptoms, viewed conversations as helpful to neutral, and came away with an increasingly positive sentiment toward the use of chatbots for proactive stress management. The follow-up interviews with a subset of the participants indicated that almost half of the common daily stressors could be discussed with chatbots, potentially reducing the burden on human coping resources. In the future, we plan to add new features such as web-based learning recommendation systems while conducting longitudinal studies on the efficacy of the Popbots to serve as an effective public health tool.

Acknowledgments

This study was supported, in part, by the Stanford Institute for Human-Centered Artificial Intelligence and Groupe Renault. The contributions by MLM were made while in transition from Stanford School of Medicine to the University of Delaware.

Conflicts of Interest

PEP is currently the founder and chief executive officer of Pop-Up Health Inc. However, the elaboration of this manuscript precedes the creation of this company.

Multimedia Appendix 1

Additional information about our formative study, examples of conversations with our chatbots, and additional figures.

[DOCX File , 191 KB - formative_v5i9e25294_app1.docx ]

References

3. Patient adoption of mHealth: use, evidence and remaining barriers to mainstream acceptance. IMS Institute for Healthcare Informatics. URL: https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/patient-adoption-of-mhealth.pdf [accessed 2021-08-26]


42. Kroenke K, Spitzer RL, Williams JB, L.

43. Telegram. Telegram Messenger. URL: https://telegram.org/

44. Kroenke K, Spitzer RL, Williams JB, L.


47. Kroenke K, Spitzer RL, Williams JB, L.


Abbreviations

AI: artificial intelligence
CBT: cognitive behavioral therapy
PHQ: Patient Health Questionnaire
WOZ: Wizard of Oz
Original Paper

Semisupervised Deep Learning Techniques for Predicting Acute Respiratory Distress Syndrome From Time-Series Clinical Data: Model Development and Validation Study

Carson Lam¹, MD; Chak Foon Tso¹, PhD; Abigail Green-Saxena¹, PhD; Emily Pellegrini¹, MEng; Zohora Iqbal¹, PhD; Daniel Evans¹, MS; Jana Hoffman¹, PhD; Jacob Calvert¹, MSc; Qingqing Mao¹, PhD; Ritankar Das¹, MSc
Dascena, Inc, Houston, TX, United States

Corresponding Author:
Carson Lam, MD
Dascena, Inc
Suite B, Private Mailbox 65148
12333 Sowden Rd
Houston, TX, 77080
United States
Phone: 1 7149326188
Email: clam@dascena.com

Abstract

Background: A high number of patients who are hospitalized with COVID-19 develop acute respiratory distress syndrome (ARDS).

Objective: In response to the need for clinical decision support tools to help manage the next pandemic during the early stages (ie, when limited labeled data are present), we developed machine learning algorithms that use semisupervised learning (SSL) techniques to predict ARDS development in general and COVID-19 populations based on limited labeled data.

Methods: SSL techniques were applied to 29,127 encounters with patients who were admitted to 7 US hospitals from May 1, 2019, to May 1, 2021. A recurrent neural network that used a time series of electronic health record data was applied to data that were collected when a patient’s peripheral oxygen saturation level fell below the normal range (<97%) to predict the subsequent development of ARDS during the remaining duration of patients’ hospital stay. Model performance was assessed with the area under the receiver operating characteristic curve and area under the precision recall curve of an external hold-out test set.

Results: For the whole data set, the median time between the first peripheral oxygen saturation measurement of <97% and subsequent respiratory failure was 21 hours. The area under the receiver operating characteristic curve for predicting subsequent ARDS development was 0.73 when the model was trained on a labeled data set of 6930 patients, 0.78 when the model was trained on the labeled data set that had been augmented with the unlabeled data set of 16,173 patients by using SSL techniques, and 0.84 when the model was trained on the entire training set of 23,103 labeled patients.

Conclusions: In the context of using time-series inpatient data and a careful model training design, unlabeled data can be used to improve the performance of machine learning models when labeled data for predicting ARDS development are scarce or expensive.

(JMIR Form Res 2021;5(9):e28028) doi:10.2196/28028

KEYWORDS

acute respiratory distress syndrome; COVID-19; semisupervised learning; deep learning; machine learning; algorithm; prediction; decision support

Introduction

Acute respiratory distress syndrome (ARDS) is a broadly defined clinical syndrome associated with significant morbidity and mortality [1,2]. ARDS has been critically misdiagnosed and underdiagnosed despite the high ARDS-associated mortality rates and high rates of related hospital resource use [2-4]. Confidence in ARDS diagnosis varies due to the heterogeneity in disease presentation [5] as well as the heterogeneity in the disease’s definition [6,7]. The identification of ARDS across clinical settings remains subjective [8], and it can be difficult
to diagnose the syndrome in patients with underlying conditions that have similar symptom presentations, such as pneumonia [9].

Early intervention is critical to improving patient outcomes, yet there remains a need for clinical decision support tools that can accurately predict ARDS development prior to onset. Per the current Berlin definition of ARDS [10], a radiology report is required to diagnose ARDS. However, rapid radiology reports are often unavailable due to a lack of access to equipment or the lack of the consideration of ARDS by clinicians [11]. The variability in ARDS presentation also makes it challenging to predict ARDS development by using standard machine learning methods, which typically require large amounts of confidently labeled data for supervised learning [12]. Semisupervised learning (SSL) paradigms have been applied to the tasks of biological data [13] classification and microRNA [14] classification and to many similar classification tasks in the domain of biotechnology [15-17] to address the dual issues of poor label quality and limited data quantity. In the context of early ARDS prediction, SSL is useful because it allows for the implicit specification of a useful gold standard. An SSL model schema that integrates information from many clinical features (including radiology reports) during training but only requires a small set of readily available clinical features to make predictions based on test data may, in practice, be crucial to improving early ARDS prediction. The aim of this study was to provide a proof of concept that SSL may be useful for predicting ARDS onset.

**Methods**

**Data Sets**

Data from 7 hospital systems were used in this study, including data from patients who were monitored in emergency department, inpatient ward, and intensive care unit settings. All data were collected passively and deidentified in compliance with the Health Insurance Portability and Accountability Act. Patients with a length of hospital stay of at least 3 hours were included, and positive encounters were defined by the gold standard described in the **Gold-Standard Labels** section. The data set was divided into hold-out test sets, training sets, validation sets, and unlabeled sets, as shown in Figure 1. In order to set aside an external hold-out test set, patients from 3 of the 7 hospital systems were considered to be a part of the test set, and there was no overlap between the patients in this test set and the patients from the remaining 4 hospital systems that were used for the validation, training, and unlabeled sets. Of the 25,670 patients from the non-test set, 2567 (10%) were set aside for the validation set. Of the remaining 23,103 nonvalidation, nontest patients, 6930 (30%) were set aside for the labeled data set, and 16,173 (70%) were set aside for the unlabeled data set. The true label of the unlabeled data set, by definition, was never revealed during the SSL process.

**Gold-Standard Labels**

A patient was defined as developing ARDS if a new diagnosis of ARDS based on International Classification of Diseases (ICD) codes appeared in the patient’s chart and if we could verify (ie, by using the physiologic time-series data) that the patient experienced respiratory failure. A new code was defined as a code that appeared after admission and was not present during the 1000 hours leading up to admission. In total, 7 outcomes were labeled for each patient, as follows:

1. A clinical diagnosis of ARDS was determined by using ICD codes. The ARDS ICD codes used were J80, J96.0, J96.2, J96.9, and 518.81.
2. Respiratory failure was defined according to the accepted criteria for respiratory failure (a peripheral oxygen saturation [SpO2] level of <92% or a partial pressure of oxygen [PaO2]/fraction of inspired oxygen [FiO2] ratio of <300) [18]. These were approximately corresponding points on the oxygen-hemoglobin dissociation curve, and they allowed us to identify the earliest possible time point in which respiratory failure occurred, even when the PaO2 level had not been measured. The prediction of ARDS

---

**Figure 1.** Sample size allocation in the data set. ARDS: acute respiratory distress syndrome; LOS: length of stay.
development leading to respiratory failure was the primary task, and the area under the receiver operating characteristic curve (AUROC) and area under the precision recall curve (AUPRC) were computed and reported based on this label. Although they were not the primary focus of this paper, secondary auxiliary outcomes were used as well.

2. A COVID-19 diagnosis was defined as a positive polymerase chain reaction test for new COVID-19 ICD codes—U07.1, B97.21, B97.29, J12.81, and B34.2.

3. Acute kidney injury was defined by using the following ICD codes: N17, N19, and R34.

4. A broad class of thrombosis was defined by using the following ICD codes: 112, I26, I63, I67, I74, 180, I81, and I82.

5. Sepsis was defined by using the following ICD codes: A40, A41, R65.2, T81.12, T81.44, O85, and O86.04.

6. Patients were labeled according to whether—a drop in SpO₂ (below 97%)—they were eventually placed on mechanical ventilation.

Onset Time

The time point for which the algorithm prediction was outputted was the first time point when the SpO₂ level fell below the lower range of normal (SpO₂<97%). This was referred to as the prediction time. The onset time for ARDS-positive encounters was defined as the first time point at which any ARDS-related ICD code was found in a patient’s electronic health record (EHR). The onset time for respiratory failure was the first time point when the SpO₂ level fell below 92% or the PaO₂/FiO₂ ratio fell below 300. To find these time points, our data processing function first analyzed all of the SpO₂ values that were measured for any given patient; if any measurements were <97%, we saved the date-time entry. After this below-97% measurement was collected, we proceeded to determine if the following two later events occurred:

1. The addition of an ARDS ICD code into the EHR. If found, the date-time entry for this event was saved, and the date-time entry for the below-97% SpO₂ event was subtracted from that of the subsequent measuring event before converting the time difference to hours and plotting the data in a histogram.

2. The subsequent measuring of an SpO₂ level of <92% or a PaO₂/FiO₂ ratio of <300. If found, the date-time entry for this event was saved, and the date-time entry for the below-97% SpO₂ event was subtracted from it before converting the time difference to hours.

Input Features

ARDS predictions were made by using a defined set of data types or features across all hospitals, regardless of the data availability at a particular hospital. Model input features were chosen based on the efficiency at which the features could be extracted from EHRs, feature availability, and consultation with clinicians. For example, most definitions of ARDS require lung findings to be present in the absence of heart failure [3]. The feature availability for the data set is presented in Figure S1 in Multimedia Appendix 1. The model input features consisted of the following: age, gender, the initiation of antibiotics prior to the prediction time, the initiation of supplemental oxygen prior to the prediction time, a history of heart failure, systolic and diastolic blood pressure, heart rate, temperature, respiratory rate, SpO₂ level (pulse oximetry), creatinine level, blood urea nitrogen level, bilirubin level, glucose level, the international normalized ratio, white blood cell count, red blood cell count, platelet count, percent neutrophil count, percent lymphocyte count, percent monocyte count, hematocrit level, lactate level, aspartate transaminase level, and alanine transaminase level. Not all features were required for the model to make a prediction of ARDS onset.

Data Processing

The time-series data were organized as a matrix with rows that represented features and columns that represented update time steps. This method of organizing time-series clinical data was the same method used by Che et al [19]. Each column represented a time step in which an update had occurred for one of the features. For simplicity, the first 6 rows represented the following constant features: age, male gender, female gender, the initiation of antibiotics prior to the prediction time, the initiation of supplemental oxygen prior to the prediction time, and a history of heart failure. Except for age, which was normalized by using the mean and SD of the training set, the remaining constant features were coded as 1 or 0. The time series features each had 2 rows—one row contained missingness masks (ie, measurements that were current for a given time step were coded as 1; otherwise, they were coded as 0), and the other row contained the normalized value of current measurements. Further, a row was used to denote the minutes that had passed since the last time step. This was normalized according to the mean and SD of the duration of time between time steps in the training set. To manage memory usage, we set a limit of 32 time steps prior to the prediction time. For patients with less than 32 time steps prior to the prediction time, we performed zero-padding and represented the resulting values as missing data by using a 0 in the missingness mask row. Details of our missing data processing methodology are presented in Table S1 in Multimedia Appendix 1.

Machine Learning Models

The recurrent neural network (RNN) was implemented with the PyTorch package (version 1.40) in Python 3.6 [20]. The demographics and time series measurements were organized into a sequence of vectors and normalized before being passed to the RNN component of the model by using a normalization layer, as follows:

$$n(v) = a \odot \left( \frac{|v - \mu|}{\sigma + \epsilon} \right) + b \ (1)$$

In equation 1, \(n(v)\) is a normalization function that learns the parameters mean (\(\mu\)), SD (\(\sigma\)), scaling factor \(a\), and translation factor \(b\) to normalize vector embeddings (\(v\)). The symbol “\(\odot\)” is the Hadamard product (also known as the element-wise product). \(\epsilon\) was set to 1e⁻⁷ to prevent division-by-zero errors.

For the RNN, a sequence module—a 2-layer gated recurrent unit (GRU) [21] with 64 hidden units—was used. A soft attention module was used to assign scores to each time step in the sequence. The attention score was a learned importance weight for each time step. This weight was converted into a point, the initiation of supplemental oxygen prior to the prediction time, a history of heart failure, systolic and diastolic blood pressure, heart rate, temperature, respiratory rate, SpO₂ level (pulse oximetry), creatinine level, blood urea nitrogen level, bilirubin level, glucose level, the international normalized ratio, white blood cell count, red blood cell count, platelet count, percent neutrophil count, percent lymphocyte count, percent monocyte count, hematocrit level, lactate level, aspartate transaminase level, and alanine transaminase level. Not all features were required for the model to make a prediction of ARDS onset.

Data Processing

The time-series data were organized as a matrix with rows that represented features and columns that represented update time steps. This method of organizing time-series clinical data was the same method used by Che et al [19]. Each column represented a time step in which an update had occurred for one of the features. For simplicity, the first 6 rows represented the following constant features: age, male gender, female gender, the initiation of antibiotics prior to the prediction time, the initiation of supplemental oxygen prior to the prediction time, and a history of heart failure. Except for age, which was normalized by using the mean and SD of the training set, the remaining constant features were coded as 1 or 0. The time series features each had 2 rows—one row contained missingness masks (ie, measurements that were current for a given time step were coded as 1; otherwise, they were coded as 0), and the other row contained the normalized value of current measurements. Further, a row was used to denote the minutes that had passed since the last time step. This was normalized according to the mean and SD of the duration of time between time steps in the training set. To manage memory usage, we set a limit of 32 time steps prior to the prediction time. For patients with less than 32 time steps prior to the prediction time, we performed zero-padding and represented the resulting values as missing data by using a 0 in the missingness mask row. Details of our missing data processing methodology are presented in Table S1 in Multimedia Appendix 1.

Machine Learning Models

The recurrent neural network (RNN) was implemented with the PyTorch package (version 1.40) in Python 3.6 [20]. The demographics and time series measurements were organized into a sequence of vectors and normalized before being passed to the RNN component of the model by using a normalization layer, as follows:

$$n(v) = a \odot \left( \frac{|v - \mu|}{\sigma + \epsilon} \right) + b \ (1)$$

In equation 1, \(n(v)\) is a normalization function that learns the parameters mean (\(\mu\)), SD (\(\sigma\)), scaling factor \(a\), and translation factor \(b\) to normalize vector embeddings (\(v\)). The symbol “\(\odot\)” is the Hadamard product (also known as the element-wise product). \(\epsilon\) was set to 1e⁻⁷ to prevent division-by-zero errors.

For the RNN, a sequence module—a 2-layer gated recurrent unit (GRU) [21] with 64 hidden units—was used. A soft attention module was used to assign scores to each time step in the sequence. The attention score was a learned importance weight for each time step. This weight was converted into a
probability distribution and multiplied by each sequence’s deepest hidden activation in the GRU to create a weighted sum of the activations, which is called the context vector. We concatenated the context vector to the final GRU embedding and passed this vector to a 2-layer feed-forward neural network to produce an output vector for classification. The output vector’s length (7 dimensions) was equal to the number of target labels. The intermediate layer before the output logits was a 64D representation of each patient, which was referred to as the penultimate embedding. Similar to the method used by Bahdanau et al [21], the score of the attention neural network was parameterized by a feed-forward neural network, as follows:

\[
\text{score}(l, h) = K \cdot \tanh (A \cdot \text{prelu}(B \cdot n([l, h])))
\]  

In equation 2, \(\tanh\) and \(\text{prelu}\) denote the hyperbolic tangent function and parameterized rectified linear unit nonlinearity functions, respectively. \(l\) denotes the last value in the sequence and the deepest hidden activation in the GRU, \(h\) denotes each and any hidden activation in the deepest layer of the GRU in the sequence, and \([l, h]\) denotes the concatenation of and into a longer vector (the length of the individual vectors were added together). \(K, A,\) and \(B\) denote the learned matrix parameters of the neural network. The symbol “\(\cdot\)" denotes matrix multiplication.

The whole GRU-RNN, attention module, and classification module were end-to-end differentiable, which allowed for optimization from input to output. The attention neural network was a mechanism of the RNN that allowed for higher quality learning. Rather than summarizing a time series of vectors, the attention neural network assigned each vector a score according to how important the vector was in terms of allowing the model to make a prediction. As such, the attention network mechanism allowed the RNN to focus on specific parts of the input, thereby improving model performance. The RNN model schema is presented in Figure 2.

**Figure 2.** RNN model schema. ARDS: acute respiratory distress syndrome; DiasBP: diastolic blood pressure; GRU: gated recurrent unit; HR: heart rate; RespRate: respiratory rate; RNN: recurrent neural network; SpO\(_2\): peripheral oxygen saturation; SysBP: systolic blood pressure; Temp: temperature.

Each point in the RNN model schema was representative of a neuron. The neurons received data input from vital signs and laboratory measurements that were recorded in EHRs. At each layer, the RNN combined information from the current and previous time points to update the activations in the deepest hidden layer of the GRU, which, when combined with the importance-weighted average generated by the attention neural network, created a summary of all time-series data—the context vector. The last layer was a feed-forward neural network, which used the activation size of the last deepest hidden state in the GRU combined with the context vector (64+64=128) as input data. With this RNN schema, the model was trained to predict...
the primary and auxiliary target labels simultaneously and to evaluate a loss function based on all targets.

Model Training

Overview

Our method of SSL was a combination and adaptation of the methodology that was previously developed by Li et al [22] and Xie et al [23]. Rather than performing whole-document and image classification, which were conducted in these prior studies, our models were designed to perform their prediction task by using multivariate time-series data. Our models were tasked with predicting ARDS onset in both the general population and patients with COVID-19.

Initial Pseudolabeling

Our methodology builds on our prior work [24]; we simplified the prediction time and inputs for the model. The RNN was first trained on the labeled training set, without making use of the unlabeled set, until convergence occurred in the validation set (keeping the model with the most minimal validation loss). The first RNN was called the pseudolabeler or initial teacher. The initial teacher was used to predict the probability of future ARDS and auxiliary target development for every patient in the unlabeled set. The mean probability was used as the threshold for the temporary label (the pseudolabel). If the initial teacher assigned a probability that was higher than twice the mean probability for that sample, the sample was considered to be positive and added to the SSL pseudolabeled data set for this cycle of training. If the initial teacher assigned a probability that was below the mean, the sample was considered to be negative and added to the SSL pseudolabeled data set. The remaining samples were not used for this cycle of training because they were considered to be “unconfident.”

Semisupervised Relabeling

An RNN was used as the semisupervised learner or student machine learning algorithm. For each cycle of SSL, the student machine learning algorithm was trained on the combined labeled and pseudolabeled training set. Afterward, it was fine-tuned on the labeled training set. The student machine learning algorithm then became the teacher for the next cycle of SSL by relabeling the pseudolabeled (unlabeled) training set. The SSL training setup was not meant to perform well on the auxiliary targets; instead, the 6 auxiliary outcomes were used as a multitasking form of regularization for the primary problem. The validation set was used for both hyperparameter selection and the prevention of overfitting only with respect to the ARDS outcome and not with respect to the other outcomes. The pseudolabeling and selection of “confident” labels for the next SSL cycle was performed only with respect to the ARDS outcome and not with respect to the other outcomes. A new RNN was initialized, and the cycle was repeated. Models were trained for 40 epochs, and the model with the best validation set performance was saved (Figure 3).

RNN training was performed by using the Adam optimizer [25] with a decay scheduler to scale down the learning rate (starting from 0.001) by a factor of 0.9 when the multiclass binary cross-entropy loss increased over 2 epochs. A batch size of 2048 was parallelized over 4 Nvidia Tesla M60 (Nvidia Corporation) graphics processing units.

Figure 3. Semisupervised learning schema. The colored network represents the initiation of a new model.

Performance Evaluation

Following SSL training, the initial teacher and student models were evaluated for their performance on a hold-out test set based on the AUROC, AUPRC, sensitivity, specificity, positive predictive value, and negative predictive value. The initial teacher performance on the test set defined the baseline performance that SSL was meant to improve upon. In addition to reporting this SSL performance, to define a ceiling for performance, we also compared SSL performance to the
performance of a model that was trained on the labeled set and unlabeled set by using the gold-standard labels for both sets instead of the pseudolabels. This model, which was trained on the nonvalidation, non-test patient data, was referred to as the *all data model*. Principal component analysis and t-stochastic neighbor embedding were used to conduct dimensionality reduction and perform a cluster analysis on the RNN’s intermediate representations.

**Results**

Demographically, patients with ARDS were similar to patients without ARDS. Except for cardiovascular disease, including heart failure, patients with ARDS had a higher incidence of chronic pulmonary disease, hypertension, diabetes, and obesity (Table 1).

The median time interval from the prediction time until the onset of ARDS, which appeared as a diagnosis in patients’ EHRs, was 59 hours. The median time interval from the prediction time until the onset of respiratory failure, which appeared as a drop in \( \frac{2}{FiO_2} \) ratio of <300, was 21 hours. Histograms of the time intervals for the whole data set are shown in Figure 4, and those for the test set are shown in Figure S2 in Multimedia Appendix 1.

<table>
<thead>
<tr>
<th>Demographic characteristics</th>
<th>Patients with ARDS(^a) (n=3383), n (%)</th>
<th>Patients without ARDS (n=74), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-30</td>
<td>93 (2.7)</td>
<td>3 (4.1)</td>
</tr>
<tr>
<td>30-39</td>
<td>131 (3.9)</td>
<td>5 (6.8)</td>
</tr>
<tr>
<td>40-49</td>
<td>156 (4.6)</td>
<td>6 (8.1)</td>
</tr>
<tr>
<td>50-59</td>
<td>373 (11)</td>
<td>8 (10.8)</td>
</tr>
<tr>
<td>60-69</td>
<td>577 (17.1)</td>
<td>14 (18.9)</td>
</tr>
<tr>
<td>≥70</td>
<td>1886 (55.7)</td>
<td>34 (45.9)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1649 (48.7)</td>
<td>36 (48.6)</td>
</tr>
<tr>
<td>Female</td>
<td>1734 (51.3)</td>
<td>38 (51.4)</td>
</tr>
<tr>
<td><strong>Race and ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>61 (1.8)</td>
<td>3 (4.1)</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>23 (0.7)</td>
<td>1 (1.4)</td>
</tr>
<tr>
<td>Non-Hispanic Asian</td>
<td>1 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>3290 (97.3)</td>
<td>70 (94.6)</td>
</tr>
<tr>
<td>Non-Hispanic other</td>
<td>2 (0.1)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Unknown race or ethnicity</td>
<td>4 (0.1)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Comorbidities</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>History of chronic pulmonary disease</td>
<td>126 (3.7)</td>
<td>9 (12.2)</td>
</tr>
<tr>
<td>History of cardiovascular disease</td>
<td>551 (16.3)</td>
<td>19 (25.7)</td>
</tr>
<tr>
<td>History of chronic heart failure</td>
<td>158 (4.7)</td>
<td>6 (8.1)</td>
</tr>
<tr>
<td>History of hypertension</td>
<td>242 (7.2)</td>
<td>14 (18.9)</td>
</tr>
<tr>
<td>History of diabetes</td>
<td>186 (5.5)</td>
<td>11 (14.9)</td>
</tr>
<tr>
<td>History of cancer</td>
<td>343 (10.1)</td>
<td>13 (17.6)</td>
</tr>
<tr>
<td>History of obesity</td>
<td>73 (2.2)</td>
<td>5 (6.8)</td>
</tr>
</tbody>
</table>

\(^a\)ARDS: acute respiratory distress syndrome.
Figure 4. Prediction look-ahead times until (A) ARDS onset and (B) respiratory failure. The time until ARDS onset is the time after admission until any care provider adds the International Classification of Diseases code for ARDS into the electronic health record. The time until respiratory failure is the time after admission until the first measurement of an SpO$_2$ level of <92% or a partial pressure of oxygen/fraction of inspired oxygen ratio of <300. These samples reflect the total data set. ARDS: acute respiratory distress syndrome; SpO$_2$: peripheral oxygen saturation.

The performance results of the initial teacher model and the semisupervised RNN model on the test data set are provided in Table 2. The best validation performance was achieved on cycle 3 of 4 during SSL training.

The results in Table 2 indicate that by using 16,173 unlabeled samples, we were able to use SSL to improve the model that was trained on the 6930 labeled samples. The amount of improvement was nontrivial compared to the performance that was possible when the model was trained on all data. The AUROCs and AUPRCs for the teacher, SSL, and all data models on the hold-out test set are presented in Figure 5. The same curves for auxiliary targets are provided in Figure S3 in Multimedia Appendix 1. Data on the subset of 489 patients with COVID-19 in the test set are shown in Figure S4 in Multimedia Appendix 1.

Table 2. Teacher and semisupervised learning model performance on test set.

<table>
<thead>
<tr>
<th>Performance indicator</th>
<th>Initial teacher model</th>
<th>Semisupervised learning model</th>
<th>All data model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area under the receiver operating characteristic curve</td>
<td>0.73</td>
<td>0.78</td>
<td>0.84</td>
</tr>
<tr>
<td>Area under the precision recall curve</td>
<td>0.035</td>
<td>0.045</td>
<td>0.065</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.76</td>
<td>0.78</td>
<td>0.78</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.55</td>
<td>0.61</td>
<td>0.72</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>0.020</td>
<td>0.023</td>
<td>0.033</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>0.995</td>
<td>0.996</td>
<td>0.996</td>
</tr>
</tbody>
</table>

The attention weights generated by the RNN were probed to visualize the signals that were attended to by the RNN. This method was used to implicitly describe the importance that was assigned to each feature by the model and provided some clues about model interpretability. For each patient in the test set, the time step with the greatest attention weight was extracted. This was the focus time step. The feature vector at this time step was interpreted as a z score for the subset of features that were measured during this particular time step. For example, a value of $-0.5$ in the heart rate dimension would denote that the heart rate is half an SD lower than the mean. For each time varying feature, we accumulated these directional inflections across all focus time steps and plotted a normalized heat map (Figure 6). Consistent with our intuition, the time steps with the greatest attention weights had large negative inflections in SpO$_2$ level and large positive inflections in respiratory rate.
**Discussion**

**Principal Findings**

We present a method of SSL for the early prediction of ARDS development. To address the challenges of poor label quality and limited data quantity, which make it difficult to predict ARDS development by using standard machine learning methods, we developed a method of SSL whereby confidently labeled data were assigned to a labeled data set and used for the testing, validation, and training of the RNN machine learning model. In the SSL scheme, the RNN model learned the latent representation of ARDS that was present in unlabeled data and expanded its own understanding of gold-standard labels. In doing so, the model established a relational link between a small set of commonly available clinical features and ARDS without needing to explicitly learn the Berlin definition of ARDS. To supplement the comparatively small labeled training data set, an unlabeled data set was pseudolabeled by an initial teacher RNN model. The pseudolabeled data were used for pretraining...
and were iteratively re-pseudolabeled by an evolving RNN-based machine learning model after the model was fine-tuned on the labeled training set. The SSL method was capable of accurately predicting ARDS development and was a considerable improvement over the baseline teacher model. Since the model was constructed by using a small subset of clinical features and outperformed a baseline model that was trained only on the small subset of labeled data, in practice, the model could be applied in settings where many clinical features are not available and settings where existing ARDS labels are incomplete or of low quality.

The paradigm described in this study differs from those in similar published machine learning studies because we apply an SSL methodology to the task of predicting the development of a severe respiratory condition (ie, a complication of COVID-19). In the case of other clinical conditions for which similar methodologies have been implemented (eg, predicting sepsis [26] and detecting microaneurysms and vascular lesions [27-29]), elements of the clinical definitions of such conditions can often be matched by using widely available EHR data. However, in the case of ARDS, measurements that can be used to create reliable gold-standard labels are not as widely available. This lack of data availability is detrimental to the supervised training of an ARDS prediction tool, as there may be many patient encounters that cannot be labeled as those involving ARDS and may in fact involve an episode of ARDS.

If we had restricted ourselves to a supervised learning approach, which has been applied in the context of other clinical prediction tasks [30-32], our options for working with unlabeled data would have been limited. Alternatively, assigning these encounters a label of non-ARDS would have undermined the interpretation of performance metrics. We were therefore motivated to apply an SSL methodology to the task of ARDS prediction not only by the potential to improve upon our prior work [24] and to address new clinically relevant applications of machine learning, but also by the need to approach ARDS prediction in a fundamentally new way to address the practical challenges associated with a lack of reliably labeled retrospective data. Importantly, the prediction tool developed in this study can be used to accurately predict ARDS development without the requirement of radiographic data or subjective interpretation.

Among general populations and COVID-19 populations in settings where radiographic information may not be available, the tool could be used to provide advance warning for ARDS onset and may allow for timely intervention. This would be particularly impactful for health care providers working in regions of lower socioeconomic status, where funding for advanced medical infrastructure and access to vaccines are limited, as these regions are known to have a higher incidence of burdens resulting from severe COVID-19 [33]. In addition, the SSL approach can leverage a small amount of costly labeled data (eg, during radiographic or manual adjudication by physicians for pseudolabeling a large amount of training data) to improve model performance.

There are several limitations to this study that lend themselves to opportunities for future work. To make the model applicable to a wide variety of clinical care settings, we simplified the model input features. Over the course of testing the SSL model, we also observed that model performance varied across clinical settings. It is possible that some hospitals may have collected features that were more important to making predictions or that features may have been collected more frequently in some hospitals than in others. In addition, most SSL methods involve some form of data augmentation in addition to pseudolabeling, and it remains an open question as to how to best perform data augmentation with clinical time-series data. In future work, we aim to determine if reinforcement learning is a suitable and mathematically rigorous methodology for the augmentation of clinical time-series data. Moreover, as we stressed earlier, predicting true ARDS development by using the Berlin definition requires radiology data. In the future, we would like to include radiology data in our model and compare the model presented in this study to the Berlin gold standard. On the other hand, our attention weight heat map (Figure 6) aims to provide insight about what signals were most attended to by the RNN.

Although it provides useful data, information such as temporal change and the waveform of signals are lost in the heat map. Finally, model performance was only assessed based on retrospective patient data, and we were therefore unable to determine how the models might perform in prospective settings. Prospective validation is required to evaluate the impact of model predictions on patient outcomes.

Conclusions

An SSL model was developed and externally validated for early ARDS prediction in both the general population and patients with COVID-19. Higher performance was achieved by the SSL model compared to that of the baseline teacher model for the general intensive care unit patient population. The semisupervised machine learning methodology allowed for early ARDS prediction in a manner that successfully mitigated the challenges that are commonly associated with a lack of reliably labeled data.

**Acknowledgments**

We gratefully acknowledge Anna Siefkas and Gina Barnes for their assistance in revising this manuscript. Funding for this study was provided by the National Science Foundation under grant number 2014829.

**Authors’ Contributions**

CL and CFT performed the data analysis for this work. CL, CFT, ZI, EP, AGS, DE, and JH contributed to the drafting of this work. All authors contributed to the revision of this work. CL, RD, JC, and QM contributed to the conception of this work.
Conflicts of Interest
All authors who are affiliated with Dascena (Houston, Texas, USA) are/were employees or contractors of Dascena. RD, JC, and QM own stock in Dascena.

Multimedia Appendix 1
Supplementary material.

References


Abbreviations

ARDS: acute respiratory distress syndrome
AUPRC: area under the precision recall curve
AUROC: area under the receiver operating characteristic curve
EHR: electronic health record
FiO_2: fraction of inspired oxygen
GRU: gated recurrent unit
ICD: International Classification of Diseases
PaO_2: partial pressure of oxygen
RNN: recurrent neural network
SpO_2: peripheral oxygen saturation
SSL: semisupervised learning

Edited by G Eysenbach; submitted 17.02.21; peer-reviewed by M Aczon, S Kim, S Shams; comments to author 31.05.21; revised version received 18.06.21; accepted 01.08.21; published 14.09.21.

Please cite as:
Semisupervised Deep Learning Techniques for Predicting Acute Respiratory Distress Syndrome From Time-Series Clinical Data: Model Development and Validation Study
JMIR Form Res 2021;5(9):e28028
URL: https://formative.jmir.org/2021/9/e28028
doi:10.2196/28028
PMID:34398784
Assessing the Care Modality Preferences and Predictors for Digital Mental Health Treatment Seekers in a Technology-Enabled Stepped Care Delivery System: Cross-sectional Study

Elissa Kozlov¹, PhD; Meghan McDarby², MA; Maximo Prescott³, MPH; Myra Altman³,4, PhD

¹Institute for Health, Health Policy and Aging Research, School of Public Health, Rutgers University, New Brunswick, NJ, United States
²Department of Psychological and Brain Sciences, Washington University in St. Louis, St. Louis, MO, United States
³Modern Health, San Francisco, CA, United States
⁴Clinical Excellence Research Center, Stanford University, Stanford, CA, United States

Corresponding Author:
Myra Altman, PhD
Modern Health
450 Sansome St, 12th Floor
San Francisco, CA, 94111
United States
Phone: 1 603 306 6231
Email: myra@joinmodernhealth.com

Abstract

Background: Access to mental health services continues to be a systemic problem in the United States and around the world owing to a variety of barriers including the limited availability of skilled providers and lack of mental health literacy among patients. Individuals seeking mental health treatment may not be aware of the multiple modalities of digital mental health care available to address their problems (eg, self-guided and group modalities, or one-to-one care with a provider). In fact, one-to-one, in-person treatment is the dominant care model with a masters- or doctoral-level trained mental health provider, and it may or may not be the appropriate or preferred level of care for an individual. Technology-enabled mental health platforms may be one way to improve access to mental health care by offering stepped care, but more research is needed to understand the care modality preferences of digital mental health care seekers because additional modalities become increasingly validated as effective treatment options.

Objective: The purpose of this study was to describe and evaluate the predictors of care modality preferences among individuals enrolled in a technology-enabled stepped mental health care platform.

Methods: This exploratory, cross-sectional study used employee data from the 2021 Modern Health database, an employer-sponsored mental health benefit that uses a technology-enabled platform to optimize digital mental health care delivery. Chi-square tests and one-way analysis of variance (ANOVA) were conducted to evaluate associations among the categorical and continuous factors of interest and the preferred care modality. Bivariate logistic regression models were constructed to estimate the odds ratios (ORs) of preferring a one-on-one versus self-guided group, or no preference for digital mental health care modalities.

Results: Data were analyzed for 3661 employees. The most common modality preference was one-on-one care (1613/3661, 44.06%). Approximately one-fourth of the digital mental health care seekers (881/3661, 24.06%) expressed a preference for pursuing self-guided care, and others (294/3661, 8.03%) expressed a preference for group care. The ORs indicated that individuals aged 45 years and above were significantly more likely to express a preference for self-guided care compared to individuals aged between 18 and 24 years (OR 2.47, 95% CI 1.70-3.59; \( P < .001 \)). Individuals screening positive for anxiety (OR 0.73, 95% CI 0.62-0.86; \( P < .001 \)) or depression (OR 0.79, 95% CI 0.66-0.95; \( P = .02 \)) were more likely to prefer one-on-one care.

Conclusions: Our findings elucidated that care modality preferences vary and are related to clinical severity factors and demographic variables among individuals seeking digital mental health care.

(JMIR Form Res 2021;5(9):e30162) doi:10.2196/30162

KEYWORDS
stepped care; technology; mental health care; patient-centered care
Introduction

Equitable access to mental health services continues to be a systemic problem in the United States and around the world [1]. Barriers to treatment for mental disorders include attitudinal barriers (eg, treatment skepticism) and structural barriers (eg, insufficient mental health workforce) [2]. An especially potent structural barrier to accessing mental health services is that prospective patients have challenges identifying and accessing viable treatment options [3], a key element of mental health literacy [4]. Importantly, individuals seeking mental health treatment may not understand the range of options available to address their problem, let alone expressing preferences for different modalities of receiving digital mental health care (eg, self-guided or group care, and one-to-one care with a provider).

In fact, the dominant care model of one-to-one, in-person treatment involving a masters- or doctoral-level trained mental health provider may or may not be the appropriate or preferred level of care for an individual. This model of care, which requires access to trained and often expensive mental health specialists, partially explains the worldwide treatment gap in mental health care, as only a fraction of individuals with mental health needs receive treatment [5,6]. As a result, health care delivery systems have attempted to develop solutions that increase patient access to a variety of care options and account for barriers to treatment such as low mental health literacy and provider shortages. These models are known as “stepped care” approaches, which attempt to match patients to care options based on symptom severity and perceived needs [7,8].

Although they are not regularly available to the general population, technology-enabled mental health platforms may be one way to improve access to mental health care [9]. These platforms have the potential to streamline and optimize mental health care by matching patients’ presenting problems, severity, and treatment modality preferences. Importantly, these platforms have the potential to create an opportunity for individuals to address their problem, let alone expressing preferences for different modalities of care.

Additionally, technology-enabled mental health platforms have the potential to optimize access to mental health services by facilitating stepped care in digital mental health treatments. Although there are many people seeking mental health care, some care seekers may not need or want traditional one-on-one psychotherapy given their presenting problem and severity level. A recent study that assessed care modality preferences found that less than half (44.5%) of patients with depression preferred in-person psychotherapy over digital mental health treatments (self-, peer- or provider-guided treatment) [10]. The stepped care approach posits that many care seekers would benefit from less resource-intensive treatments such as self-guided or group-based digital mental health treatments, which are more scalable than individual, in-person psychotherapy treatments.

For certain populations with subclinical symptoms or areas of concern outside of traditional psychopathology, there may be no supporting evidence or need for individual psychotherapy from expensive and difficult-to-find specialists.

As more evidence-based modalities of receiving digital mental health care emerge— including self-guided interventions delivered via the internet or mobile health (mHealth) technology [11,12], group-based videoconferencing [13], and video-delivered individual psychotherapy sessions with a provider [14]—it is essential to better describe and understand the predictors of patient preferences for these modalities of digital mental health care. Prior research has demonstrated that individuals express preferences for mental health care when asked, and when those preferences are not met, the psychological outcomes are affected [15]. Existing clinical guidelines also encourage providers to incorporate patient preferences when evaluating treatment options wherever possible [16]. To facilitate patient-centered stepped care, more research is needed to understand care modality preferences because additional digital mental health treatments are becoming increasingly validated as effective options.

The purpose of this exploratory cross-sectional analysis of existing data was to examine modality preferences among individuals seeking digital mental health treatments through a technology-enabled, stepped care platform. We analyzed data from employees who registered with Modern Health, an employer-sponsored mental health benefit that uses a technology-enabled platform to optimize mental health care delivery. Our aim was to describe the care modality preferences of digital mental health treatment seekers and evaluate the associations among demographic factors, clinical factors, and the primary reasons for seeking care. We hypothesized that digital mental health care seekers with demographic characteristics traditionally associated with fewer treatment-seeking behaviors, such as being older (40 years and above) and being males, would be more likely to state a preference for self-guided care rather than traditional one-on-one treatment. We also hypothesized that individuals with higher levels of clinical severity would be more likely to state a preference for one-on-one care.

Methods

Intervention

Modern Health utilizes a stepped care approach to mental health care by directing users to the appropriate level of care when initiating treatment. All users answer a series of questions during registration to determine if their care needs correspond to preventive care, moderate clinical care, or high clinical care. The platform assesses clinical needs as well as each user’s care modality preferences to tailor treatment recommendations. Given that the study period coincided with the COVID-19 pandemic, only digital mental health treatments were available to the users of the platform and included the following:

https://formative.jmir.org/2021/9/e30162
self-guided digital courses, group support via videoconferencing, one-to-one telecoaching with in-app texting, and one-to-one teletherapy (video-delivered individual psychotherapy sessions) with a licensed mental health specialist. The self-guided digital courses include guided meditations and modules that cover topics such as cognitive behavioral therapy, stress management, resilience and coping, burnout, and establishing healthy habits.

Participants
The participants were employees (N=3661) who registered to use a mental health benefits platform between February 18, 2021, and April 9, 2021, and had provided complete registration data. Because Modern Health gradually rolled out the registration assessment, participants with missing data do not reflect poor responses but rather differences in when the registration portal was updated for different users. We analyzed data from individuals who were 18 years or older, had access to a smartphone, tablet, or computer, completed all baseline assessment questions through the Modern Health platform, and had their demographic data recorded. This study was reviewed by the WIRB-Copernicus Group Institutional Review Board (WCG IRB) and determined to be exempt from Institutional Review Board oversight.

Procedures
Eligible employees register for Modern Health using a mobile app or via a website. Upon registering, participants complete a baseline assessment that includes the World Health Organization-5 Well-being Index (WHO-5), Patient Health Questionnaire-2 (PHQ-2), Generalized Anxiety Disorder 2-item (GAD-2) questionnaire, and a questionnaire about their primary focus areas and their care modality preference.

Measures

Demographics
Employers optionally provided the gender and age data for employees eligible to use the Modern Health benefit prior to registration.

Well-being
Well-being was assessed using the WHO-5, a robust and unidimensional assessment of subjective well-being that has high psychometric validity as well as adequate sensitivity and specificity to screen for depressive symptoms [17]. Scores range on the percentage scale from 0 to 100 with higher scores indicating greater well-being.

Depression
The PHQ-2 was used to screen for depression. The PHQ-2 asks individuals if they have been feeling down, depressed, or hopeless and if they have had little interest or pleasure in doing things. The score totals range from 0 to 6 and cutoff scores higher than 3 are considered a positive screen for depression. In a recent study of community-based participants, the PHQ-2 showed a sensitivity of .64 and specificity of .85, which were comparable to the longer version of the scale, the PHQ-9 [18].

Anxiety
The GAD-2 was used to screen for anxiety. The GAD-2 is a psychometrically robust screener for anxiety that asks participants if they have been feeling nervous, anxious, or on the edge and if they have had difficulties in being able to stop or control worrying. The total scores on the GAD-2 range from 0 to 6, with scores higher than 3 indicating a positive screen for a clinically significant anxiety disorder. In a recent study of community-based participants, the GAD-2 showed a sensitivity of .71 and a specificity of .69, which were comparable to the longer version of the scale, the GAD-7 [18].

Topic Selection
The topics that participants selected during onboarding as their reason for visiting the platform were organized by their corresponding well-being dimensions (“my emotions,” “my physical well-being,” “my relationships,” and “my finances”). They selected these from a pre-established list of potential topics, such as anxiety, depression or low mood, improving my relationships and communication, burnout, and general professional development. The participants could not enter their own topics; they had to choose from the pre-established list.

Functional Impairment
An item adapted from the WHO Short Disability Assessment was used to assess functional impairment. Participants were asked, “In the past 2 weeks, (topic selections) have made it difficult for me to function in my life at home and work.” The response options followed a Likert scale including “strongly agree,” “agree,” “neither disagree or agree,” “disagree,” and “strongly disagree.”

Care Modality Preferences
Care modality preferences were assessed for individuals seeking digital mental health treatments on the platform by asking, “When it comes to improving my mental health, I prefer to work:…” Participants were able to select a single answer from the following response options: “on my own (self-guided, at my own pace),” “with a small group (live community sessions led by care professionals),” “one-on-one (meet with a care professional),” or “I’m not sure.”

Statistical Analysis
Data cleaning and analysis was performed using R (version 4.0.3), a statistical software. WHO-5 scores were mean-centered and scaled to improve interpretability during regression modeling, such that a value of 0 represents the mean and an increase of 1 unit represents a difference of 1 SD. A complete case analysis was performed such that an individual’s data were only included if registration was completed and demographic data were available. Descriptive statistics were used to describe the demographic, clinical, and primary reasons for seeking care, and care modality preference characteristics of the sample. Chi-square and Kruskal-Wallis tests were used to evaluate associations between the categorical and continuous factors of interest and the preferred care modality, respectively. Bivariate logistic regression models were constructed to estimate the odds ratios (ORs) describing the relative differences in the odds of selecting self-guided or group modalities or being unsure of

https://formative.jmir.org/2021/9/e30162

JMIR Form Res 2021 | vol. 5 | iss. 9 | e30162 | p.107
(page number not for citation purposes)
modality preferences compared to the odds of a preference for one-on-one care within each factor of interest.

**Results**

**Descriptive Data**

The mean age of respondents was 35.2 years (SD 9.4; range 19-74). The sample comprised mostly females (2113/3661, 57.7%). Respondents reported mean well-being scores of 43.33 (IQR 28; range 0-100), which can be interpreted as reduced well-being according to a commonly used cutoff score of 50 [17]. The primary topic selection endorsed most frequently by respondents was “my emotions” (1772/3661, 48.4%), followed by “my professional life” (707/3661, 19.3%), “my relationships” (560/3661, 15.3%), “my physical well-being” (549/3661, 15%), and lastly, “my finances” (73/3661, 2%). Approximately 35% of the sample (1271/3661) screened positive for anxiety, and 22.4% of the respondents (819/3661) screened positive for depression. The most selected care modality preference was traditional one-on-one care (1613/3661, 44.06%). Approximately one-fourth of the respondents (881/3661, 24.06%) expressed a preference for obtaining self-guided care. Less than 10% of the respondents (294/3661, 8.03%) reported a preference for small-group care options, whereas nearly a quarter of the respondents (873/3661, 23.85%) were unsure of their preferred treatment modality. The demographic, clinical, and topic selection characteristics differed significantly across care modality preferences. Table 1 presents the descriptive data.
Table 1. Descriptive statistics and associations between preferred care modalities and demographic, clinical, and primary reasons for seeking care.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Care modality preference</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total (N=3661)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>On my own (self-guided) (n=881, 24.06%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>One-on-one care (meet with a care professional) (n=1613, 44.06%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>With a small group (live community sessions led by care professionals) (n=294, 8.03%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I'm not sure (n=873, 23.85%)</td>
<td>&lt;.001^a</td>
</tr>
<tr>
<td>Age (years), n (%)</td>
<td>18-24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>243 (6.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>58 (6.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>119 (7.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15 (5.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>51 (5.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25-34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1900 (51.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>397 (45.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>925 (57.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>141 (48)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>437 (50.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35-44</td>
<td></td>
</tr>
<tr>
<td></td>
<td>895 (24.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>214 (24.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>393 (24.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>75 (25.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>213 (24.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>45+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>623 (17)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>212 (24.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>176 (10.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>63 (21.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>172 (19.7)</td>
<td></td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td>Female</td>
<td>&lt;.001^a</td>
</tr>
<tr>
<td></td>
<td>2113 (57.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>443 (50.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>955 (59.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>181 (61.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>534 (61.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1548 (42.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>438 (49.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>658 (40.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>113 (38.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>339 (38.8)</td>
<td></td>
</tr>
<tr>
<td>Subjective well-being (WHO-5^b score)</td>
<td>Mean</td>
<td>&lt;.001^c</td>
</tr>
<tr>
<td></td>
<td>43.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>47.56</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40.63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>47.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>42.81</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IQR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Median (range)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>44 (0-100)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>48 (0-100)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40 (0-100)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>44 (0-92)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40 (0-100)</td>
<td></td>
</tr>
<tr>
<td>PHQ-2^d screening result, n (%)</td>
<td>Negative depression screen (score &lt;3)</td>
<td>&lt;.001^a</td>
</tr>
<tr>
<td></td>
<td>2842 (77.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>762 (86.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1172 (72.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>236 (80.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>672 (77)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive depression screen (score ≥3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>819 (22.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>119 (13.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>441 (27.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>58 (19.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>201 (23)</td>
<td></td>
</tr>
<tr>
<td>GAD-2^e screening result, n (%)</td>
<td>Negative anxiety screen (score &lt;3)</td>
<td>&lt;.001^a</td>
</tr>
<tr>
<td></td>
<td>2390 (65.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>703 (79.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>911 (56.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>217 (73.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>559 (64)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive anxiety screen (score &gt;3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1271 (34.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>178 (20.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>702 (43.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>77 (26.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>314 (36)</td>
<td></td>
</tr>
<tr>
<td>Functional impairment, n (%)</td>
<td>Strongly agree</td>
<td>&lt;.001^a</td>
</tr>
<tr>
<td></td>
<td>455 (12.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>57 (6.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>288 (17.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>22 (7.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>88 (10.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agree</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1536 (42)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>308 (35)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>726 (45)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>121 (41.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>381 (43.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neither disagree nor agree</td>
<td></td>
</tr>
<tr>
<td></td>
<td>864 (23.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>231 (26.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>329 (20.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>74 (25.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>230 (26.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Disagree</td>
<td></td>
</tr>
<tr>
<td></td>
<td>579 (15.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200 (22.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200 (12.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>54 (18.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>125 (14.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Strongly disagree</td>
<td></td>
</tr>
<tr>
<td></td>
<td>227 (6.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>85 (9.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>70 (4.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>23 (7.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>49 (5.6)</td>
<td></td>
</tr>
<tr>
<td>Primary focus area, n (%)</td>
<td>My emotions</td>
<td>&lt;.001^a</td>
</tr>
<tr>
<td></td>
<td>1772 (48.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>327 (37.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>896 (55.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>129 (43.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>420 (48.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>My finances</td>
<td></td>
</tr>
<tr>
<td></td>
<td>73 (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>26 (3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>27 (1.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7 (2.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13 (1.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>My physical well-being</td>
<td></td>
</tr>
<tr>
<td></td>
<td>549 (15)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>242 (27.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>104 (6.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>59 (20.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>144 (16.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>My professional life</td>
<td></td>
</tr>
<tr>
<td></td>
<td>707 (19.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200 (22.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>284 (17.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>62 (21.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>161 (18.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>My relationships</td>
<td></td>
</tr>
<tr>
<td></td>
<td>560 (15.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>86 (9.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>302 (18.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>37 (12.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>135 (15.5)</td>
<td></td>
</tr>
</tbody>
</table>

^aPearson chi-square test.
^bWHO-5: World Health Organization-5 Well-being Index.
^cKruskal-Wallis test.
^dPHQ-2: Patient Health Questionnaire-2.
^eGAD-2: Generalized Anxiety Disorder-2.
Associations Between Demographic Characteristics and Care Modality Preferences

The results of the multinomial logistic regression analysis are presented in Table 2. Preferring a self-guided care modality over one-on-one care with a provider was significantly associated with older age, being males, higher well-being, screening negative for anxiety or depression, and reporting less functional impairment. The ORs indicated that individuals aged 45 and above were significantly more likely to prefer self-guided care over one-on-one care compared to individuals aged between 18 and 24 (OR 2.47, 95% CI 1.70-3.59; P<.001). Respondents identifying themselves as males were also significantly more likely to prefer self-guided care (OR 1.43, 95% CI 1.22-1.69; P<.001). More reports of well-being predicted a preference for self-guided care (OR 1.45, 95% CI 1.34-1.58; P<.001). Individuals who screened positive for anxiety (OR 0.33, 95% CI 0.27-0.40; P<.001) or depression (OR 0.42, 95% CI 0.33-0.52; P<.001) were significantly less likely to prefer self-guided care. The likelihood of preferring self-guided care was significantly lower among individuals who neither agreed nor disagreed (OR 0.58, 95% CI 0.40-0.83; P=.003), agreed (OR 0.35, 95% CI 0.25-0.49; P<.001), or strongly agreed (OR 0.16, 95% CI 0.11-0.25; P<.001) that their topic selection caused functional impairment. In addition, individuals who selected “my finances” (OR 2.64, 95% CI 1.52-4.59; P=.001), “my physical well-being” (OR 6.38, 95% CI 4.9-8.29; P<.001), or “my professional life” (OR 1.93, 95% CI 1.55-2.41; P<.001) as their topic were significantly more likely to prefer a self-guided modality compared to individuals who reported “my emotions” as a primary area of focus.

A preference for a group care modality over one-on-one care with a provider was significantly associated with older age, higher well-being, screening negative for anxiety or depression, and reporting less functional impairment. Respondents aged 45 and above were significantly more likely to prefer group care (OR 2.84, 95% CI 1.54-5.22; P<.001). More reports of well-being also predicted a preference for group care (OR 1.41, 95% CI 1.24-1.60; P<.001). Individuals who screened positive for anxiety (OR 0.46, 95% CI 0.35-0.61; P<.001) or depression (OR 0.65, 95% CI 0.48-0.89; P=.007) were significantly less likely to prefer group care. The likelihood of preferring group care over one-on-one care was significantly lower among individuals who agreed (OR 0.51, 95% CI 0.30-0.84; P=.009) or strongly agreed (OR 0.23, 95% CI 0.12-0.44; P<.001) that their topic selection had resulted in functional impairment. Respondents who indicated that “my physical well-being” was their primary area of focus were significantly more likely to prefer a group care modality (OR 3.94, 95% CI 2.72-5.70; P<.001), as was the case for respondents who reported that “my professional life” was their primary area of focus (OR 1.52, 95% CI 1.09-2.11; P=.014).

Being unsure about one’s preference for treatment over one-on-one care with a provider was significantly associated with older age, greater well-being, screening negative for anxiety or depression, and reporting less functional impairment. Individuals aged over 45 years were more likely to be unsure about their treatment modality preference (OR 2.28, 95% CI 1.54-3.37; P<.001). Individuals reporting higher well-being were significantly more likely to be unsure about their treatment modality preference (OR 1.13, 95% CI 1.04-1.23; P=.005). Individuals who reported “my physical well-being” as their primary topic were significantly more likely to report that they were unsure about their treatment modality preferences (OR 2.95, 95% CI 2.24-3.90; P<.001).

Screening positive for anxiety (OR 0.73, 95% CI 0.62-0.86; P<.001) or depression (OR 0.79, 95% CI 0.66-0.95; P=.019) was significantly associated with a preference for one-on-one care. The likelihood of preferring one-on-one care was significantly higher among individuals who strongly agreed that their topic selection had caused functional impairment (OR 0.44, 95% CI 0.28-0.68; P<.001).
Table 2. Comparison of care modality preferences based on bivariate multinomial logistic regression results for relative associations between preferred care modalities and demographic, clinical, and primary reasons for seeking treatment.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Care modality preference distribution</th>
<th>Self-guided vs 1:1 with provider</th>
<th>Group vs 1:1 with provider</th>
<th>Unsure vs 1:1 with provider</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1:1 (Ref)</td>
<td>Self-guided (%)</td>
<td>Group (%)</td>
<td>Unsure (%)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24 (ref)</td>
<td>49</td>
<td>23.9</td>
<td>6.2</td>
<td>21</td>
</tr>
<tr>
<td>25-34</td>
<td>48.7</td>
<td>20.9</td>
<td>7.4</td>
<td>23</td>
</tr>
<tr>
<td>35-44</td>
<td>43.9</td>
<td>23.9</td>
<td>8.4</td>
<td>23.8</td>
</tr>
<tr>
<td>45+</td>
<td>28.3</td>
<td>34</td>
<td>10.1</td>
<td>27.6</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (ref)</td>
<td>45.2</td>
<td>21</td>
<td>8.6</td>
<td>25.3</td>
</tr>
<tr>
<td>Male</td>
<td>42.5</td>
<td>28.3</td>
<td>7.3</td>
<td>21.9</td>
</tr>
<tr>
<td>Subjective well-being</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WHO-5 score&lt;sup&gt;d&lt;/sup&gt;</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Depression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative PHQ-2&lt;sup&gt;e&lt;/sup&gt;  screen (ref)</td>
<td>41.2</td>
<td>26.8</td>
<td>8.3</td>
<td>23.6</td>
</tr>
<tr>
<td>Positive PHQ-2 screen</td>
<td>53.8</td>
<td>14.5</td>
<td>7.1</td>
<td>24.5</td>
</tr>
<tr>
<td>Anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative GAD-2&lt;sup&gt;f&lt;/sup&gt;  screen (ref)</td>
<td>38.1</td>
<td>29.4</td>
<td>9.1</td>
<td>23.4</td>
</tr>
<tr>
<td>Positive GAD-2 screen</td>
<td>55.2</td>
<td>14</td>
<td>6.1</td>
<td>24.7</td>
</tr>
<tr>
<td>Functional impairment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly agree</td>
<td>63.3</td>
<td>12.5</td>
<td>4.8</td>
<td>19.3</td>
</tr>
<tr>
<td>Agree</td>
<td>47.3</td>
<td>20.1</td>
<td>7.9</td>
<td>24.8</td>
</tr>
<tr>
<td>Neither agree nor disagree</td>
<td>38.1</td>
<td>26.7</td>
<td>8.6</td>
<td>26.6</td>
</tr>
<tr>
<td>Disagree</td>
<td>34.5</td>
<td>34.5</td>
<td>9.3</td>
<td>21.6</td>
</tr>
<tr>
<td>Strongly disagree (ref)</td>
<td>30.8</td>
<td>37.4</td>
<td>10.1</td>
<td>21.6</td>
</tr>
<tr>
<td>Primary focus area</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My emotions (ref)</td>
<td>50.6</td>
<td>18.5</td>
<td>7.3</td>
<td>23.7</td>
</tr>
<tr>
<td>My finances</td>
<td>37</td>
<td>35.6</td>
<td>9.6</td>
<td>17.8</td>
</tr>
<tr>
<td>My physical well-being</td>
<td>18.9</td>
<td>44.1</td>
<td>10.7</td>
<td>26.2</td>
</tr>
<tr>
<td>My professional life</td>
<td>40.2</td>
<td>28.3</td>
<td>8.8</td>
<td>22.8</td>
</tr>
</tbody>
</table>

https://formative.jmir.org/2021/9/e30162
understand patient factors associated with different modality preferences. Although the largest number of participants (45%) expressed a preference for traditional one-on-one treatment, nearly half of the participants indicated a preference for self-directed care (24%) or being unsure of their preference (24%). Older age, being males, lower overall distress, and negative depression and anxiety screening results were significantly predictive of a preference for self-guided digital care. This suggests that many adults would prefer a self-guided digital approach to manage the challenges associated with subclinical psychological distress. This aligns with prior research that men and middle-aged and older-aged adults tend to seek less help for psychological distress [20,21]. Notably, our study revealed that 25% of the participants were unsure regarding their care modality preference. Participants unsure about their preference were more likely to be older than 45 years with lower overall distress, and negative depression and anxiety screening results. These results indicate that more psychoeducation about care modalities may be warranted for up to a quarter of care-seeking individuals to help patients self-determine their care preferences. Given that our sample included only individuals whose employers offered the Modern Health mental health benefit, it may be reasonable to assume that this is a particularly well-educated and well-resourced population. These results are likely to be more exaggerated in the general population that tends to have less mental health access and literacy. Future research should investigate the relationship between mental health literacy and prior experience with mental health care with perceived needs and preferences for different types of mental health care.

Technology-enabled mental health care delivery systems, though not commonly available to the general public, have the potential to approach psychological care in a way that is patient-centered and individualized to patient preferences and needs for treatment. Importantly, technology-enabled mental health platforms have the ability to ensure that patient care is collaborative between patients and providers and that patient values guide clinical decisions [22]. This study revealed that traditional one-on-one mental health care, which is frequently regarded as the “gold standard,” may not be preferred to the same extent across patients. When presented with the opportunity to choose, some patients prefer group care, self-directed treatment, or care options that are less rigidly structured (eg, meeting with a coach when needed instead of...
once a week/every week for months, meeting for 30-minute sessions, having check-ins once a month, and support via text messaging). A technology-enabled platform can customize care options based on preferences and perceived needs.

Limitations
The sample of respondents in the current study is a relatively homogenous group of primarily younger adults having access to the Modern Health mental health benefit through their employer. Similarly, these respondents are likely to be generally healthier, better educated, and more financially stable than the general population, given their affiliation with the Modern Health employer-based benefit. As a result, these findings may not be generalizable to a more diverse sample. Future research should seek to confirm these results in a community-based sample with greater heterogeneity in the respondent characteristics. Another limitation of this study is that certain demographic variables were not collected (e.g., race, ethnicity, and income); thus, our ability to completely characterize the sample was hindered. Future research can build upon this study to more comprehensively characterize the demographic variables associated with care modality preferences. Additionally, this study did not enquire participants about their previous experiences with mental health care, which is a factor likely to inform treatment modality preferences and mental health literacy. Future studies can seek to understand additional factors that influence patient preferences for mental health treatment modalities.

Conclusions
This study revealed that care modality preferences for digital mental health treatment are variable based on demographic factors as well as clinical severity and area of focus indicators. This suggests that care modality preferences align with the innovations in mental health care delivery; one-on-one care with a provider is no longer the only or necessarily best option for many care seekers, as internet-delivered group and self-paced interventions have also shown strong clinical effectiveness for certain populations [14,23,24]. To provide efficient, scalable, and patient-centered mental health care, it is essential to continue understanding how best to funnel care seekers into different treatment modalities within a stepped care model. Our study revealed several key clinical and demographic factors that were associated with different care preferences, but future research should investigate how other important patient-level factors—including mental health literacy, race, ethnicity, and prior experience with the mental health care systems—impact care modality preferences and how aligning care recommendations with modality preferences affects care usage and treatment outcomes.

Acknowledgments
We would like to acknowledge Danielle H Llaneza and Sara Sagui-Henson for their assistance with formatting and editing the manuscript.

Conflicts of Interest
MP and MA are employees of Modern Health and have equity in the company. The other authors report no conflicts of interest.

References


Abbreviations

ANOVA: analysis of variance
CBT: cognitive behavioral therapy
GAD-2: Generalized Anxiety Disorder-2
mHealth: mobile health
OR: odds ratio
PHQ-2: Patient Health Questionnaire-2
WHO-5: World Health Organization-5 Well-being Index (WHO-5)

Edited by G Eysenbach; submitted 04.05.21; peer-reviewed by E Olsson, M Feijt; comments to author 22.05.21; revised version received 21.06.21; accepted 01.08.21; published 15.09.21.

Please cite as:
Kozlov E, McDarby M, Prescott M, Altman M
Assessing the Care Modality Preferences and Predictors for Digital Mental Health Treatment Seekers in a Technology-Enabled Stepped Care Delivery System: Cross-sectional Study
JMIR Form Res 2021;5(9):e30162
URL: https://formative.jmir.org/2021/9/e30162
PMID: 34439665
Original Paper

Designing User-Centered Mobile Health Initiatives to Promote Healthy Behaviors for Children With Disabilities: Development and Usability Study

Keiko Shikako¹, BSc, PhD; Ebele R I Mogo¹, PhD; Valerie Grand-Maison², MPH; Robert Simpson³, MSc; Lesley Pritchard-Wiart⁴, PhD; Annette Majnemer¹, PhD; Jooay App Research Group⁵

¹School of Physical and Occupational Therapy, McGill University, Montreal, QC, Canada
²Guelph University, Montreal, QC, Canada
³MAB-Mackay Centre School, Montreal, QC, Canada
⁴Department of Physical Therapy, University of Alberta, Edmonton, AB, Canada
⁵McGill University, Montreal, QC, Canada

Corresponding Author:
Keiko Shikako, BSc, PhD
School of Physical and Occupational Therapy
McGill University
3654 Promenade Sir William Osler
Montreal, QC, H3G 1Y5
Canada
Phone: 1 514 398 4400
Email: keiko.thomas@mcgill.ca

Abstract

Background: The gap between research and its practical application in community settings limits its impact on public health. Closing this gap has the potential to improve the well-being of underserved groups, such as children with disabilities. Mobile health has the potential to improve access to community resources and support for underserved populations, thereby encouraging improved health behaviors.

Objective: In this feasibility pilot study, we describe the development of the mobile app Jooay. Jooay was developed in partnership with stakeholders to facilitate access to leisure and physical activity community programs for children and youth with disabilities. We also reflect on the lessons learned throughout the implementation process that are relevant for improving the health behaviors of children with disabilities.

Methods: We used a participatory action research approach to develop the app. We also administered a survey to current Jooay users and analyzed various app usage indicators to explore use patterns, user feedback, and preferences. Finally, we critically appraised the implementation process through a best practices for implementation research framework.

Results: We developed a product that responds to users’ identified need to find information and follows accessibility and user-centered design standards. The analysis of usage data revealed that access to the Jooay app is concentrated in urban areas. Perceptions, attitudes, and information needs varied according to the type of user. The use of the mobile app changed over time, and usage decreased after the app was downloaded, indicating a need for the sustained engagement of app users. Users found value in the ability to identify activities that they would not otherwise know about. However, app use alone was not sufficient to improve participation. Although the app was developed based on users’ active input in multiple iterations, we encountered challenges with survey recruitment and attrition, suggesting the need for more seamless and engaging means for collecting data within this population.

Conclusions: Interactions between users and the app can sustain user engagement and behavior change. We will improve the app’s next iterations by using the information gained from this study to conduct a larger study to assess the relationship among social and material deprivation, urban design, and access to inclusive and adaptive leisure programs. This study will inform the improvement of app listings to improve the use of Jooay by different user groups and promote health through mobile apps for marginalized groups.
implementation research; mobile health; children with disabilities; physical activity promotion; digital health; inclusive leisure participation; mobile phone

Introduction

Background

Implementation research involves studying research uptake and its effect on the outcomes of multiple stakeholders [1,2]. In medicine and public health, a significant gap remains between research knowledge and action, limiting the direct impact of research on health [1]. Closing this gap requires the consideration of multiple contextual variables, consideration of technology use, and other strategies to improve health behaviors [1].

Innovative strategies for facilitating changes in health behavior are particularly important for underserved communities and populations. People with disabilities comprise approximately 15% of the global population and are at risk of poor health outcomes [3]. Children and youth with disabilities face challenges in accessing health services and health-promoting activities [4]. They rarely meet the recommended standards for physical activity and have lower physical activity levels than their peers [5-8]. They also have a higher prevalence of noncommunicable diseases, such as obesity, diabetes mellitus, and coronary artery disease [8], and are disproportionately affected by environmental, socioeconomic, and interpersonal barriers to healthy lifestyles [8]. Although parents or other caregivers (hereafter caregivers) value their children’s participation in physical activities, they face multiple participation challenges, such as inadequate access to adapted programs and inclusive settings [9-11].

Health promotion initiatives can improve outcomes for children living with a diverse range of social, emotional, and behavioral disabilities [12-14]. Contextual factors that can serve as participation barriers or facilitators for children with disabilities include information about activities, the cost of activities, the accessibility of facilities, and the presence of trained staff and support [3,15-19]. The alignment of needs of people with disabilities with effective health promotion initiatives can foster better health outcomes, provide a sense of empowerment, reduce health disparities, and improve overall individual quality of life and community well-being [15].

Mobile health (mHealth) is gaining primacy for the creation of targeted, accessible, and context-appropriate health promotion solutions. mHealth tools include various devices, software, and solutions that use mobile phones to improve health [20]. Potential benefits include time savings, convenience, and improved access to underserved populations [21-23]. mHealth tools have also improved health behaviors among young people and in chronic disease management [14,24,25].

Although there have been several pilot studies on mHealth interventions, knowledge gaps remain regarding the appropriate development and use of mHealth to promote health equitably [22]. Preliminary studies suggest that users of mHealth are younger, more educated, have better health, and belong to a higher income group than nonusers [26]. Despite its promise in improving health, mHealth may actually exacerbate health disparities if underserved populations do not have access to digital tools and the ability to develop literacy in using them.

In addition, research on mHealth has not yet sufficiently demonstrated efficacy, effectiveness, user engagement, effective scale-up, and competitive value [21,27]. Many mHealth efforts have been inadequately designed for implementation and evaluation [21]. Evaluations that do exist do not prioritize data disaggregation, limiting considerations of equity and impact for marginalized groups [21,22].

Objective

In this study, we explore the feasibility of mHealth to improve access to information on community-based inclusive leisure activities for children with disabilities. Our specific objectives are to (1) describe the development process of the mHealth solution and (2) identify use patterns and user preferences. The secondary objective is to establish the feasibility of using a mobile app to test behavior change and to pilot test data collection through app analytics and users.

Understanding the development, implementation, and uptake of mHealth can facilitate the design and use of future technologies for health promotion in high-risk groups. The knowledge gained will also inform the scale-up of the app and, more broadly, the field of implementation research.

Methods

Overview

In this pilot feasibility study [28], we asked the following question: can a mobile app be developed in collaboration with stakeholders and used to promote health behavior changes in children with disabilities, and if so, how? (feasibility component). We also conducted a pilot study to assess the extent to which the app supports health behavior change (pilot component). In this paper, we describe the app development process and the results of a small-scale survey with a subset of app users.

We adopted a hybrid implementation research design [29], whereby the intervention or solution is developed and tested concurrently instead of using the traditional approach, in which development is conducted before the intervention is tested with the population comes first and then is tested out in the population followed by testing. The value of hybrid designs resides in the possibility of cocreating knowledge while simultaneously incorporating and testing intervention improvements. Multiple iterations required in a technology development project make the hybrid design an optimal approach to effectively test and implement user-responsive mHealth-based interventions.
Ethical approval for this study was obtained from the McGill University Institutional Ethics Review Board as well as the Centre for Interdisciplinary Research in Rehabilitation.

Figure 1. Integration of theoretical frameworks. mHealth: mobile health.

**Theoretical Frameworks**
Several theoretical frameworks informed our approach and study objectives. Figure 1 illustrates the integration of these frameworks with the project objectives and procedures.

**The overarching objective and the development process (specific objective 1)** were informed by the best practices for implementation research prescribed by Peters et al [31]. Specifically, we used the following steps: (1) identify the audience and how they will use the research; (2) clearly describe the program, practice, or policy being implemented; (3) examine the implementation strategy thoroughly; (4) describe the real-world context and sample population clearly; and (5) appropriately consider outcome variables of the implementation.

The last 2 steps consider specific context variables and appropriately consider changes in contextual factors over time, unintended consequences, and system complexity.

To expand on step 5 (appropriately considering outcome variables of the implementation), we adopted the Health Behavior Framework [32]. This framework includes a detailed consideration of outcome variables related to the expected implementation (ie, use of mHealth technology, the Jooay app). The elements of the framework informed the key aspects to consider while identifying users’ patterns and preferences (specific objective 2). Determination of the feasibility of use requires a comprehensive understanding of the variables involved in the health behavior change proposed by this framework.
makers, and grassroots organizations) and were asked to consider barriers to access to leisure participation. Subsequently, they were grouped into mixed groups representing different stakeholders to discuss implementation solutions to promote the participation of children with disabilities.

Each participant was invited to write their selected top solution on a card. Instructions were that the top solution depicted in the card should be actionable, be feasible, and have an impact on the participation levels of children with disabilities. Participants then engaged in a prioritization exercise in which they exchanged cards and, in pairs, compared and ranked 2 ideas. The solution that ranked higher at the end of 5 rounds of paired ranking was considered the one that, according to the forum participants, would yield the highest impact on participation in leisure for children with disabilities. The solution ranked as the most feasible and promising solution to overcome systems barriers and promote participation in leisure across the 4 forums was “[a]n electronic list of inclusive and adapted activities.” Participants in each forum volunteered to form a working group to support the development of the idea.

Advisory Group
The Jooay app was developed in 2 phases: (1) prototype development and (2) test version development.

A convenience sample of 5 users (2 caregivers, 1 physical educator, and 2 occupational therapists [OTs]) and several research team members (including community organization leaders and physical and recreational therapists) provided input into the user interface development, user experience development, and final test versions during both phases.

A larger group of stakeholders comprising clinicians (OTs and physical therapists), physical educators, caregivers of children with disabilities, and representatives of community organizations constituted the user partners who provided input during the development phase of the app. These stakeholders provided insights into the best features and types of information to be listed and were involved in the conception of the app, the testing of multiple versions, and the development of the research protocols to address the feasibility and pilot testing of the app. The stakeholders were chosen based on the following considerations: (1) willingness to participate; (2) direct experience working or advocating for children with disabilities; and (3) representation across diverse areas of involvement that included pediatric rehabilitation centers, school boards, community organizations, newsletters of organizations, and networks related to adapted leisure and addressing caregivers of children with disabilities.

A dynamic protocol for testing and responses was established using a collaboration platform and the research team made connections between the developers and user testers.

Prototyping and Test Versions Development
User Experience
The mobile app was initially developed as a prototype during a hackathon (weekend event grouping developers, user experience designers, programmers, and project managers). The first test version was developed for iOS, Android, and the web.

It was made available to users free of charge through regular channels (App Store, Google Play, and website).

Development
A partnership with an external mobile app developer was necessary to secure further development, and multiple funding sources were required to attain industry standards. Suggestions that arose from user partners and technology developers resulted in the creation of 3 native platforms for iOS, Android, and the web.

Core technology developments included mandatory implementation of full accessibility protocols for mobile and web platforms (including voice-over, voice control, color contrast, and easy access—a feature available in the iOS and Android accessibility protocols where buttons and number of clicks to action are reduced). An accessibility consultant was hired to test and assess these features during development.

Content
The participants provided important information on preferences for the type of information displayed for each activity listed. The selection of included domains had an impact on development cost; therefore, there was an assessment of the most relevant information components to be retained in the app.

The final domains reflected stakeholders’ preferences for information and included activity description, types of equipment required, type of disability (eg, physical, intellectual, or those classified as all are welcome), cost, and time frame. The same decision algorithm was used to create filters within the map search as well as to collect basic demographic information from users when they registered to use the app.

Stakeholders also suggested a list of other resources to be listed on Jooay, such as reference links to other types of supports toward leisure participation, such as respite care and support groups for youth and caregivers, research related to leisure, and the Jooay web-based Facebook community. Our user partners supported the development of the survey questions by reviewing the initial questions and providing feedback on how to phrase the questions, question content, and structure (eg, options for participants to suggest other fields or useful resources that should be added to the app). They also supported survey distribution to a larger sample of community organizations and rehabilitation centers, in addition to the list of app users, and acted as champions to disseminate the survey and the app as a product to other users.

Activities in the app were initially populated on the basis of pre-existing lists of adapted and inclusive leisure activities from pediatric rehabilitation centers in the targeted provinces and schools serving children with disabilities in 6 provincial capitals across Canada. Using the key terms identified on the websites of these organizations, the research team searched for additional activities and continued populating the app database.

Data Collection and Sample
This pilot project used 2 sources of data: (1) analytics of the Jooay app users and (2) electronic surveys sent to registered participants.
Participants

Through a preliminary analysis of app analytics data from approximately 600 users, we determined that approximately half were health or education service providers and the other half were caregivers of children with disabilities. No further sociodemographic information was required from the users upon registration. Furthermore, 2 different surveys were subsequently developed to target service providers and caregivers. This pilot study was conducted with all app users who had registered with the app using their email. Registration was not a mandatory requirement to gain access to information on Jooay.

A total of 273 of approximately 600 user emails were initially available. Additional users who downloaded the app and registered during the 6-month data collection period also received an invitation to participate. A brief explanation of the study and survey links were also posted through social media channels, specifically through parent support groups and the Jooay page on Facebook, asking app users who might not have registered emails to complete the survey. Study knowledge brokers in pediatric rehabilitation centers, school boards, community organizations, newsletters of organizations, and networks related to adapted leisure and addressing caregivers of children with disabilities also shared information about the app and survey. Additional participants in the regions where the activities were published on Jooay at the time of the study, namely, the 6 provinces—Alberta, British Columbia, Quebec, Nova Scotia, Saskatchewan, and Ontario—were targeted in addition to existing users of the app.

Participants were required to own a smartphone, have access to the internet, and understand English or French. Users aged <13 years were excluded from the survey, given that 13 was the established minimum age to own social media accounts, such as Facebook and Twitter, and the legal age required to download Jooay from the App Store. Assent was required for participants aged <18 years.

Procedures

Potential participants received an email with a brief explanation of the study and a link to a web-based survey on REDCap (Research Electronic Data Capture; developed by Vanderbilt University) hosted in a research database. Participants were prompted to provide consent before the initiation of the survey. A total of 5 invitations were sent by email within a 2-week interval.

Confidentiality and anonymity of survey responses were ensured by deidentifying survey responses. Participant emails were only used to prompt their participation in the study and were not associated with their answers. A 2-step password-protected REDCap account accessible only to the survey team was used to ensure data privacy. The survey questions analyzed as part of this pilot study are provided in Textbox 1.
Textbox 1. Survey questions analyzed as part of this study.

<table>
<thead>
<tr>
<th>Questions about the app and app use</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Rank the relevance of the current existing features of the app (are these sections important in helping you find an activity to pursue?)</td>
</tr>
<tr>
<td>2. Was information in the following sections helpful in finding activities?</td>
</tr>
<tr>
<td>- cost</td>
</tr>
<tr>
<td>- type of activity</td>
</tr>
<tr>
<td>- description of activity</td>
</tr>
<tr>
<td>- type of disability</td>
</tr>
<tr>
<td>- location</td>
</tr>
<tr>
<td>- age range</td>
</tr>
<tr>
<td>- schedule</td>
</tr>
<tr>
<td>- season</td>
</tr>
<tr>
<td>- reviews or ratings</td>
</tr>
<tr>
<td>- other links and related information</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Questions with &quot;agree,&quot; &quot;neutral,&quot; or &quot;disagree&quot; responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The app is easy to use</td>
</tr>
<tr>
<td>2. The app has a comprehensive list of existing activities</td>
</tr>
<tr>
<td>3. The information on the app is accurate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How did you learn about Jooay?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Social media</td>
</tr>
<tr>
<td>2. Health or education professional</td>
</tr>
<tr>
<td>3. Other sources (which one?)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Open-ended questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What feature you would like to see in the app that is currently not available?</td>
</tr>
<tr>
<td>2. What are main issues of the app?</td>
</tr>
<tr>
<td>3. Do you intend to uninstall the app? (yes or no)</td>
</tr>
<tr>
<td>If yes, explain why (open ended)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Questions about health behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The person for whom the app information is being used (child, client, or student) has engaged in regular leisure activities before using the app? (yes or no)</td>
</tr>
<tr>
<td>2. My child, client, or student participation in leisure has increased because of information found in the app? (not at all, a little-moderately, or a lot)</td>
</tr>
<tr>
<td>If participation improved, explain why (open ended)</td>
</tr>
<tr>
<td>3. The activities in the app fits my, my child’s, or my client’s needs (yes—moderately—no)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sociodemographic questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Your age (respondent)</td>
</tr>
<tr>
<td>2. Age of the person with disabilities for whom you’re using the app info</td>
</tr>
<tr>
<td>3. Province of residence</td>
</tr>
</tbody>
</table>

Analysis

After the REDCap surveys were distributed [34], responses were exported into SPSS (IBM Inc), and responses from the English and French surveys were merged. Descriptive analyses were conducted for participant characteristics, app usage patterns, and preferences for app features and information. We also explored the associations between the sociodemographic characteristics of caregivers (age and gender), with perceptions of app features, app use patterns, and other characteristics.

The question of power calculation is often a challenge in assessing the impact of mHealth tools on health outcomes [35], given the challenges of generating a significant sample size and
Reducing attrition. This pilot study also aimed to set parameters for sample size calculations for future research using this mHealth tool.

**Results**

The results of the development process and feasibility testing are presented in the context of the 2 guiding theoretical frameworks: the Implementation Framework and the Health Behavior Framework. We identify which domains of these frameworks are addressed in each of the following sections.

**Development Process of the mHealth Solution**

*Implementation Framework: Identifying the audience and how they will use the research*

*Health Behavior Framework: Health policy environment, health or community care system, community capacity and engagement, and social advocacy*

More than 200 stakeholders representing diverse groups participated in the development process of the mHealth solution. An average of 50 participants in each of the 4 stakeholder forums identified that a mobile listing of leisure activities was the most desirable and feasible solution to promote participation in leisure activities for children with disabilities.

Table 1 describes the main steps of development, the strategies adopted, and the main outputs in each step. Figure 2 shows screenshots of the final app that was developed.

Table 1. Development process and outputs.

<table>
<thead>
<tr>
<th>Steps of development</th>
<th>Strategy</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Needs assessment</td>
<td>Stakeholder forum</td>
<td>Mobile app with dynamic and interactive list of leisure activities in the community, based on geolocation (close to where children live)</td>
</tr>
<tr>
<td>Design and prototyping</td>
<td>Hackathon</td>
<td>Branding to represent multiple disability groups; English and French languages</td>
</tr>
<tr>
<td>User interface</td>
<td>Stakeholder advisory</td>
<td>Accessibility features beyond basic protocols</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Minimum information requirements upon registration</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Additional information asked for research from users (eg, sociodemographics)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Additional information given to users (eg, research about leisure and respite care)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Domains</td>
</tr>
<tr>
<td>User experience</td>
<td>Stakeholder advisory</td>
<td>Accessibility features include visual impairment, cognitive impairment, testing of map functions, and multiple platform accessibility features (iOS, Android, and web)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Easy access to key information by different users: parents versus service providers</td>
</tr>
<tr>
<td>Test versions</td>
<td>Stakeholder advisory</td>
<td>Sustainable ways to provide feedback from users to developers (email)</td>
</tr>
<tr>
<td></td>
<td>Collaboration platform (Trello; developed by Atlassian)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Third part development company</td>
<td></td>
</tr>
<tr>
<td>Pilot version</td>
<td>Public at large</td>
<td>Need to create community among users (eg, chat or group interactions)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sustainability: maintenance of updated information is crucial; maintenance of technology in each of the native platforms (cost)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Crowdsourcing: make possible for organizations and users to suggest activities</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Troubleshooting: need for ongoing technology support to maintain the app relevant users’ satisfaction-expected health outcomes</td>
</tr>
</tbody>
</table>
Stakeholders in the forums represented a range of health and policy environments, communities, and health care systems. They presented the key factors for the use of the app in the health and community environments (eg, information that should be added to make the mHealth solution relevant). Stakeholders in the forums and in the advisory group also contributed the listings that they currently had in the municipal, local listings of activities to the database and engaged in social advocacy.

**User Patterns and Preferences**

*Implementation Framework: Examining the implementation strategy thoroughly, describing the sample population clearly*

**Health Behavior Framework individual variables:** knowledge, health beliefs, social norms and supports, cultural factors and beliefs, barriers and supports, and structural factors

A total of 239 participants received the survey. The response rate was 38.9% (93/239). **Table 2** includes participant characteristics. Twenty-four participants responded to the question about age of the person for whom they were seeking activities. From those 62% (15/24) of participants indicated that they were seeking information for children aged 4-12 years.

**Table 2.** Survey participants' characteristics (N=93).

<table>
<thead>
<tr>
<th>Sample characteristics</th>
<th>Participant, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caregivers of children and youth with disabilities</td>
<td>38 (41)</td>
</tr>
<tr>
<td>Health care or education service providers</td>
<td>20 (22)</td>
</tr>
<tr>
<td>Youth with disabilities</td>
<td>4 (4)</td>
</tr>
<tr>
<td>Other stakeholder groups</td>
<td>31 (67)</td>
</tr>
<tr>
<td><strong>Living in urban area</strong></td>
<td></td>
</tr>
<tr>
<td>Caregivers</td>
<td>31 (82)</td>
</tr>
<tr>
<td>Service providers</td>
<td>18 (92)</td>
</tr>
<tr>
<td><strong>Gender (female)</strong></td>
<td></td>
</tr>
<tr>
<td>Caregivers</td>
<td>34 (89)</td>
</tr>
<tr>
<td>Service providers</td>
<td>15 (76)</td>
</tr>
</tbody>
</table>
Participants were asked to rank the relevance of app features. Participants reported that the most useful information was age range and location, followed by the activity type and description of activity. Most participants found that information in all sections of the app was helpful; sections that had less information, such as reviews and ratings (which is expected because the app is new and not many reviews had been done yet), were perceived as less helpful (Figure 3). The links provided in the settings sections included research summaries and were among the least useful sections, along with activity schedule, probably due to the frequent change in schedules, making the information not accurate.

Figure 3. Perception of users about the helpfulness of app sections.

Participants were asked to indicate (open-ended questions) the features they would like to see added to the app. A total of 34 participants responded to this question, and their responses were categorized as follows: development of a web-based community or forum (13/34, 38%), a means to track their participation in leisure and physical activities (8/34, 23%), positive prompts for action (7/34, 20%), and a points or rewards system to incentivize participation (6/34, 18%).

When asked about their concerns with the app, the main negative feedback was related to the insufficiency of listings (9/26, 35%), followed by the lack of activities in the regions where participants worked or lived (4/26, 15%). Half of the respondents agreed that the app was easy to use and accessible. However, 37% (10/26) disagreed that it was comprehensive, and 62.5% (16/26) indicated that they were neutral about whether the information was accurate. Although 85% (17/20) of the service providers using the app noted that they did not intend to uninstall the app, 72% (27/38) of the caregivers said they might intend to.

We asked participants to identify how they heard about the Jooay app as a means of increasing our understanding of how information spreads in this population. Overall, 22% (20/93) of respondents indicated social media as their source of information, followed by those who indicated that they heard about it from their physical therapist, OT, or recreational therapist (18/93, 20%); 32% (30/93) indicated other sources, such as word of mouth or through advertising in hospitals.

Respondents had the opportunity to respond to open-ended questions on whether the app contributed to increased child leisure participation, and if so, how. One of the ways in which participants indicated that the app contributed to increased participation in leisure was by raising their awareness of community resources and programs. Service providers (clinicians and educators) who responded to the survey noted that they had used the app to provide information to families, caregivers, and clients about leisure opportunities, but they could not indicate whether the app had contributed to an actual increase in leisure participation. Caregivers indicated that although the app increased their awareness of activities that they were not aware of, it was still challenging to find activities that were suitable for their children. They indicated that the app needed to include more activities to enhance opportunities for participation.
Associations

We originally intended to explore the associations of sociodemographic characteristics, such as age, gender, type of disability, and place of residence, with perceptions of the app’s features and app use patterns. However, a significant amount of missing data (61%-91% of data are missing for some variables) precluded us from doing so. However, the preliminary exploratory analysis identified significant associations between the age of the caregiver and whether the person with a disability had engaged in physical activity before using the Jooya app (n=16; Fisher-Freeman-Halton Exact test: P=.01). Most of the caregivers indicated that their child was actively engaging in physical activities before accessing the Jooya app (28/38, 75%).

There was also a significant relationship among the users who indicated that the app had the information they were looking for and the province where they were located (n=20; Fisher-Freeman-Halton Exact test: P=.05). Most of the responses (12/20, 60%) were from Quebec, and of those responses, 9 (75%) respondents found that the app had a moderate fit with their needs. The lack of diversity in the users’ responses precluded a valid test of the association between user type (caregiver vs professional) and whether users found the activities in the app were a good fit; however, half of the respondents in both categories (n=10) indicated that the activities they found in the app had a moderate to no fit with what they were looking for.

There was insufficient data from multiple user types to test for an association between user type and whether participation of their child increased as a result of the use of the Jooay app. Most of the respondents to this category of questions were caregivers (19/20, 94%), and 84% (17/20) of them indicated that their child’s participation increased after app use only slightly or moderately.

There was also an association between user type (caregiver vs service provider) and how frequently they used the app (n=24; Fisher-Freeman-Halton Exact test: P=.005). Most of the respondents to this category of questions were community organizations (7/24, 29%) and health care professionals (9/24, 37%); 45% (11/24) of respondents reported using the app multiple times without finding an appropriate activity for their client or child. The other 55% (13/24) of respondents had used the app less frequently and did not indicate whether they found an activity they were looking for or were only browsing through activities or exploring the app.

Discussion

Principal Findings

Using a PAR approach, we codeveloped and pilot-tested an mHealth tool with stakeholders to improve access to information and participation in inclusive leisure activities for Canadian children and youth with disabilities. Our aims were to understand feasibility aspects related to the use of a stakeholder-driven mHealth solution and understand user preferences and patterns to inform efforts to improve participation in leisure for children and youth with disabilities. We used the Health Behavior Framework [32] to explore the individual and social characteristics of users and their relationship with the intention to use the app, measured as downloading the app, as well as the relationship with actual behavior change, reflected in increased participation. This enabled us to explore the potential impact of the app through users’ perceptions of usefulness, preferences for specific content, and suggestions for modifications. We also began to explore the sociodemographic characteristics of users and how they relate with use of mHealth technology and the expected health behavior outcomes.

We applied the implementation framework by Peters et al [31] to understand the process-based factors that shaped the development and use of the app and opportunities to improve them. Below, we discuss the lessons learned, challenges faced, and implications for larger-scale efforts to use mHealth technology to promote health for children and youth with disabilities.

Process Development Challenges and Opportunities

Participatory research is valuable and has potential key outcomes in health and implementation science [36]. Important challenges to consider in co-designing technology with multiple stakeholders include the ethical challenges of developing study protocols that are constantly changing and require multiple ethics review board amendments, the need to respond to divergent opinions in all steps of a project (eg, questionnaires and administration forms) and product development (eg, user interface vs user experiments design phases), and the extended length of time necessary for an authentic co-design process. Most of these issues have been identified in previous reports on participatory research; however, additional challenges learned in this study include communication barriers between end user stakeholders and the technology development team because of divergent language, culture, and operational modes.

Ethics

We maintained a close discussion about the nature of the project with the institutional ethics review board and agreed on the elements that did not require ethics approval (ie, the stakeholder forums and app development stages) and which elements did (the surveys sent to app users and the information taken from app analytics). For elements requiring ethics approval, we agreed to an open protocol with the core elements of the project being initially approved through the regular, extended ethics review board procedures, and future iterations (ie, length of activity description, recruitment materials, or wording changes requested by our user collaborators) to undergo an expedited review, allowing for a reduced turnover time.

Stakeholder Engagement and Co-Design

It is important to advance implementation science on mHealth; it requires careful consideration of the interaction between technology development, participatory research and stakeholders involvement. We must develop protocols and standard operational procedures detailing aspects such as legal agreements between industry and research partners and business development plans that include design and maintenance discussions, establish a clear communication platform and verify
with end users that they are able to access it. Flexibility in accepting other forms of communication that may be preferred by stakeholders have to be considered, and detailed note taking, with designated communication contacts on the team, is ideal.

**Technology Development, User Preferences, and Health Behavior Change**

The vast majority of the app users found that the information contained in the app was relevant, albeit not comprehensive (ie, not enough activities listed in their region of residence), to really affect the desired health behavior change of increased participation for their child. This may have been the reason why 72% of caregivers planned to uninstall the app. Service providers indicated that they did not intend to uninstall the app; thus, they likely perceived greater utility. Several studies have shown that lack of information is one of the key barriers to participation and perception of good health services for families of children and youth with disabilities [37-39]. Creating a mobile app that has accurate, up-to-date information and responds to the users’ needs and preferences is a challenge but is also essential to health behavior change.

The lack of comprehensive information about activities in the app can be attributed to 2 main reasons: (1) the scarcity of activities offered in the community for this population and (2) the limited capacity to generate a comprehensive list of existing activities manually. The first issue is being addressed in a separate study (E Mogo, K Shikako, and A Majener; unpublished data; June 2021) where we conducted an in-depth analysis of the sociodemographic characteristics of regions and the availability of inclusive leisure activities, as listed in the app, with the objective of informing policy and program creation.

The second challenge relates to technology development. Creating comprehensive listings of activities that are constantly changing is a key challenge that can be addressed through technology but requires extensive sustainability planning. Sustainability and business models of mobile technology may be typical in design and industry but are foreign in health research when primary funds for development and co-design are obtained through research grants of limited duration. A sustainability model for mHealth needs to be further developed and tested to ensure the efficient use of research and user resources [40-42].

**Population Characteristics and mHealth Characteristics**

Most survey respondents were caregivers of children with disabilities living in urban areas. Most activities listed in the app are located in urban areas. In the stakeholder forums, participants indicated that it is paramount to consider mHealth solutions that target populations that face multiple layers of marginalization, such as those who live in rural areas, indigenous children and families, and those who may not have access to mobile technology. Caregivers and service providers in rural areas have been identified as populations lacking access to services and other resources [43,44]. Therefore, the utility of Jooay as an mHealth tool in these regions is limited. We intend to apply these data to inform policy on gaps in service provision.

This pilot study sheds light on the challenges of including proxies as the main users of interventions. The target population using the app are caregivers and service providers, but the actual expected behavior change (participation in leisure) is targeted at the children under their care. Information on the participation patterns of families came mainly from caregivers providing answers to the survey. We identified an association between previous participation levels and the caregivers’ age and between their familiarity with apps and the technology associated with the actual frequency of use of the Jooay app. It is known that caregiver behavior regarding leisure has an influence on the child’s level of participation [45]. The primary respondents of the survey were female (83/93, 89%), indicating that health promotion efforts could target female caregivers to affect the health behaviors of children with disabilities.

Another important characteristic of this particular mHealth solution was accessibility for persons with disabilities. Although the app users are not necessarily children and youth with disabilities (only 8/93, 9% of our sample were persons with disabilities themselves) but rather the caregivers, it is important to consider that a mobile app for persons with disabilities should comply with accessibility standards. The challenges of following accessibility standards were perceived by our accessibility consultant and were outlined in previous research [46]. It was clear from the multiple iterations of testing that industry accessibility standards are not fully accessible for different individual needs, a factor that will be considered in future app development and iterations.

**mHealth and Health Behavior Outcomes**

The survey results suggested that participants were not sure of the impact of the app on their children or clients’ participation levels but that they were certain that their knowledge about existing community activities had increased because of the app use. Ideally, mHealth should include artificial intelligence to directly track participation and objectively quantify the increase in participation as the desired outcome [47]. Although missing data prevented measuring desired behavior changes, a side effect of app use noted by some participants was the building community. Although the use of web-based communities by caregivers of patients with chronic health conditions is a relatively new phenomenon, several benefits and challenges have been identified [48]. Perceived benefits include connecting with others with similar lived experiences and challenges and increasing awareness about a medical condition or, in the present case, about existing activities and resources. Participants in this study indicated that they may be using the app not just as a resource to change health behaviors but also as a resource to connect to others and increase their awareness about possible activities, even if they are not available in the region where they live. Such indirect positive outcomes are worth investigating further. Public health implementation efforts should consider the power of connecting people and the possibilities brought about by mHealth technologies on this front.

For effective mHealth implementation, movement beyond pilot studies is needed to better understand the characteristics, preferences, and real-time use patterns of users. Partnership development with community organizations, cities, and other
layers of governments is also needed to identify solutions to link resource databases and create machine learning algorithms that maintain relevant information for the public.

**Limitations and Future Directions**

This pilot feasibility study faced limitations that supported several important considerations for future studies. First, this study was conducted in a real-world setting. We had no control over the location, type of activity offered, and the match between these activities and the participants’ preferences or needs. Previous research has indicated that preference for certain types of activities is associated with engagement in these activities [49]. This pilot study shows that, in fact, preferences are not easily matched to the availability of resources in a community, and this is a barrier to participation. This study sheds light on the importance of adapting individualized mHealth interventions to public health impacts.

The implementation strategy for this project was built in partnership with stakeholders. Implementation strategies included word of mouth and the use of local and web-based champions to disseminate information about the app and invite people to download it and use it. This project informed the important aspects of the implementation strategy on a larger scale. Recent studies have assessed the implementation of different data collection strategies through mobile apps and have reported mixed results. One study found that improving compliance with medication through digital data entry was feasible and reliable in a population of adults with HIV/AIDS [50]. Alternatively, another study that compared 3 electronic data collection methods for patients with a urinary tract infection—mobile app, electronic survey, or text message [51]—found no differences in response rates. They concluded that participants often stopped data completion after their first interaction with technology, leading to missing data. They also raise the issue of the variability of user demographics as a factor influencing response rates and preferences. We found similar challenges in completion rates. Our survey had a low response rate of approximately 39.83% and missing data, which limited the ability to make generalizations. Our stakeholder advisory group confirmed that individuals are highly interested in using the app but will respond only to very short surveys. We noticed through backend data that people often stopped completing the survey at the point at which they had to scroll through and sign the consent form before completing the survey, which also suggests the need to review ethics procedures when using mHealth technologies to conduct research.

Improvements in the recruitment strategy that will be implemented in the next phase of this study are the use of push notifications directly through the app, the shortening of the consent form to the minimal requirements, and the shortening of the survey. We will also ask our parent-partners, clinicians, and other coinvestigators to design a message that will be sent through push notifications. This message should be more welcoming of the participants’ responses. Future implementation efforts include making registration mandatory to use the app, increasing the relevance of information by implementing machine learning to update information, and increasing opportunities for interactions (ie, through push notifications and gamification). Future research must also address potential ethical constraints by appropriately adjusting their study design to elicit participation by using a mobile app as the only source for data collection.

The limited capacity of using app analytics across platforms and automated database updates imposed limitations to the data collected. Artificial intelligence to directly track participation, to objectively quantify increases in participation as the desired outcome, and improved analytics protocols are necessary to evaluate the effects on health behavior [47]. Therefore, monetary commitments associated with these strategies must be considered. Indeed, a thorough cost-effectiveness analysis should be an integral part of the scale-up implementation efforts of any technology [52].

Large-scale implementation research efforts using mHealth will need to consider better ways to engage the ecosystem of stakeholders, from users to rehabilitation centers, and community-based leisure centers to scale up and test more complex interventions. Efforts to elicit information from this population may require automated modes for data collection and partnerships between municipalities and organizations to link databases of activities and the app.

**Conclusions**

Implementation research holds the potential to drive real increments in public health by translating research findings to real-world testing. At-risk and underserved populations, such as people with disabilities, require increased efforts toward change, as they face multiple contextual barriers often leading to poor health outcomes [3].

This study piloted the development and use of Jooay, a mobile app listing inclusive leisure activities. We also sought to understand user demographics and characteristics and the corresponding variables that would be of value to support its scale-up and effectiveness testing. Our intended use of this information was to better meet the needs of children and young people with disabilities and their support systems while also informing the literature to guide similar efforts.

mHealth is promising as a viable and feasible tool to execute implementation efforts, especially for this subpopulation. mHealth tools should integrate health promotion strategies for children with disabilities considering how to overcome poverty [3], be enjoyable [19], improve access to care [15], sensitize their health care providers, be person-centered, and provide the needed support for them to engage in healthy lifestyles [3,10,15,17,18]. These tools will also have to be supported by more information to support their efficacy, effectiveness, cost utility, and engagement. Scale-up studies are necessary to move mHealth development science beyond pilot studies [21,27]. Finally, information on the demographics and characteristics of mHealth users and the impact of mHealth on behavioral predictors and health behaviors is needed [26]. This can happen if only such mHealth efforts are designed for implementation and evaluation [21].

The next phase of this project will also inform programs and policy changes that can support a sustained model of inclusive leisure activities, mHealth integration into macrosystems of
information sharing, and equitable distribution of and their families. health-promoting opportunities for children with disabilities

Acknowledgments

The Jooay Research Team is composed by multiple stakeholders who actively contribute to app and research development, including Mehrnoosh Movahed (Research Associate and Project Coordinator), Nick Katalifos (parent-partner), Joanne Charron (parent-partner), and Rachel Martens, (parent-partner and stakeholder engagement strategist); Melanie Berghorson and Stephanie Glegg (clinical partners); Jess Silver (patient-partner); Michelle McClure and Dave Sora (community partners); and Tiitu Poldma and Janet McCabe (academic collaborators). The authors would like to thank Jose Correa from McGill University for statistical consultation, Richard Cooper from McGill University for proofreading the manuscript, and Luca Martial and Icoquih Badillo for research assistance. Infrastructure support was provided by the Center for Interdisciplinary Research in Rehabilitation of the Greater Montreal hosting the Participation and Knowledge Translation in Childhood Disabilities lab, which is where the project was developed. This project was funded by the Government of Canada through a Community Accessibility Grant of the Employment and Social Development Canada, the Canadian Institutes of Health Research through the Strategic Patient-Oriented Research program CHILD-BRIGHT (Child Health Initiatives Limiting Disability - Brain Research Improving Growth and Health Trajectories) Network in partnership with the Operation Enfant Soleil and the Montreal Children’s Hospital Foundation, and partially by the Kids Brain Health Network (Network of Centres of Excellence).

Conflicts of Interest

None declared.

References


31. Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. Republished research: implementation research: what is it and how to do it: implementation research is a growing but not well understood field of health research that can contribute to more effective public health and clinical policies and programmes. This article provides a broad definition of implementation research and outlines key principles for how to do it. Br J Sports Med 2014 Apr;48(8):731-736. [doi: 10.1136/bmj.f6753] [Medline: 24659611]


34. REDCap. URL: https://www.project-redcap.org/ [accessed 2020-05-28]


Abbreviations

mHealth: mobile health

OT: occupational therapist

PAR: participatory action research

REDCap: Research Electronic Data Capture

Edited by G Eysenbach; submitted 26.08.20; peer-reviewed by C Cheng, K Arboroud-Nicotipoulos; comments to author 06.10.20; revised version received 29.01.21; accepted 19.05.21; published 16.09.21.

Please cite as:
Designing User-Centered Mobile Health Initiatives to Promote Healthy Behaviors for Children With Disabilities: Development and Usability Study
JMIR Form Res 2021;5(9):e23877
URL: https://formative.jmir.org/2021/9/e23877
doi:10.2196/23877
PMID:34528886
Electronic Video Consent to Power Precision Research: A Pilot Cohort Study

Arash Naeim1*, MD, PhD; Sarah Dry2*, MD; David Elashoff2, PhD; Zhuoer Xie3, MD; Antonia Petrus4, MBA; Clara Magyar3, PhD; Lilliana Johansen4, MPH; Gabriela Werre4, MSHSD; Clara Lajonchere5, PhD; Neil Wenger2, MD, MPH

1UCLA Center for SMART Health, Clinical Translational Science Institute, David Geffen School of Medicine at UCLA, Los Angeles, CA, United States
2David Geffen UCLA School of Medicine, Los Angeles, CA, United States
3Mayo Clinic, Rochester, MN, United States
4Office of Clinical Research, Clinical and Translational Science Institute, David Geffen School of Medicine at UCLA, Los Angeles, CA, United States
5Institute for Precision Health, David Geffen School of Medicine at UCLA, Los Angeles, CA, United States
*these authors contributed equally

Corresponding Author:
Arash Naeim, MD, PhD
UCLA Center for SMART Health
Clinical Translational Science Institute
David Geffen School of Medicine at UCLA
10911 Weyburn Ave
Los Angeles, CA, 90095
United States
Phone: 1 3107948118
Email: arashnaeim@gmail.com

Abstract

Background: Developing innovative, efficient, and institutionally scalable biospecimen consent for remnant tissue that meets the National Institutes of Health consent guidelines for genomic and molecular analysis is essential for precision medicine efforts in cancer.

Objective: This study aims to pilot-test an electronic video consent that individuals could complete largely on their own.

Methods: The University of California, Los Angeles developed a video consenting approach designed to be comprehensive yet fast (around 5 minutes) for providing universal consent for remnant biospecimen collection for research. The approach was piloted in 175 patients who were coming in for routine services in laboratory medicine, radiology, oncology, and hospital admissions. The pilot yielded 164 completed postconsent surveys. The pilot assessed the usefulness, ease, and trustworthiness of the video consent. In addition, we explored drivers for opting in or opting out.

Results: The pilot demonstrated that the electronic video consent was well received by patients, with high scores for usefulness, ease, and trustworthiness even among patients that opted out of participation. The revised more animated video pilot test in phase 2 was better received in terms of ease of use (P=.005) and the ability to understand the information (P<.001). There were significant differences between those who opted in and opted out in their beliefs concerning the usefulness of tissue, trusting researchers, the importance of contributing to science, and privacy risk (P<.001). The results showed that “I trust researchers to use leftover biological specimens to promote the public’s health” and “Sharing a biological sample for research is safe because of the privacy protections in place” discriminated opt-in statuses were the strongest predictors (both areas under the curve were 0.88). Privacy concerns seemed universal in individuals who opted out.

Conclusions: Efforts to better educate the community may be needed to help overcome some of the barriers in engaging individuals to participate in precision health initiatives.

(JMIR Form Res 2021;5(9):e29123) doi:10.2196/29123
KEYWORDS
biobanking; precision medicine; electronic consent; privacy; pilot study; video; consent; precision; innovation; efficient; precision medicine; cancer; education; barrier; engagement; participation

Introduction
Informed consent for biospecimens is an essential component for a robust program in precision medicine (PM). The use of deidentified remnant (leftover) biospecimens has come under recent scrutiny. Although the Notice of Proposed Rule Making (NPRM) to Human Subject Federal Regulations (common rule) [1] considers such tissue as not “human subjects” research, the National Institutes of Health (NIH) Genomic Data Sharing Policy expects informed consent for future research use and broad data sharing to be obtained even if the cell lines or clinical specimens are deidentified [2] (see Textbox 1 for a summary of key components of a broad consent for biospecimens). Moreover, there are many advocates and ethicists who feel there is an obligation to communicate that tissue may be used for research and to obtain informed consent [3]. Patients also want the opportunity to have their preferences dictate the use of clinical specimens for research [4].

Textbox 1. Elements of board consent.

<table>
<thead>
<tr>
<th>General requirements of study-specific informed consent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Obtaining the legally effective informed consent of the participant or the participant’s legally authorized representative</td>
</tr>
<tr>
<td>2. Seeking informed consent under circumstances that provide an opportunity to discuss and consider whether or not to participate and that minimize the possibility of coercion or undue influence</td>
</tr>
<tr>
<td>3. Providing information in understandable language</td>
</tr>
<tr>
<td>4. Providing information that a reasonable person would want to have to make an informed decision about whether to participate and providing an opportunity to discuss that information</td>
</tr>
<tr>
<td>5. Avoiding exculpatory language: Exculpatory language either waives or appears to waive the participant’s legal rights or it releases or appears to release the investigator, the sponsor, the institution, or its agents from liability for negligence.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Basic elements of study-specific informed consent</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. A description of any reasonably foreseeable risks or discomforts to the participant</td>
</tr>
<tr>
<td>7. A description of any benefits to the participant or to others that may reasonably be expected from the research</td>
</tr>
<tr>
<td>8. A statement describing the extent, if any, to which confidentiality of records identifying the participant will be maintained</td>
</tr>
<tr>
<td>9. A statement that participation is voluntary and that the participant may choose not to participate or discontinue participation at any time without penalty or loss of benefits to which the participant is otherwise entitled</td>
</tr>
<tr>
<td>10. A statement that the participant’s biospecimens—even if identifiers are removed—may be used for commercial profit</td>
</tr>
<tr>
<td>11. A statement about whether the participant will or will not share in the profit</td>
</tr>
<tr>
<td>12. A statement indicating if the research will or might include whole genome sequencing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unique elements of a broad universal consent for biospecimens</th>
</tr>
</thead>
<tbody>
<tr>
<td>13. A statement describing the types of research that may be conducted, and the information must be sufficient for a reasonable person to conclude that he or she would consent to the types of research anticipated</td>
</tr>
<tr>
<td>14. A statement describing if possible future research could raise particularly sensitive ethical, moral, religious, or cultural issues, in addition to a statement that advises the participant of the possibility that he or she might have chosen not to consent to some of those specific research studies that will use the biospecimens</td>
</tr>
<tr>
<td>15. A statement describing the identifiable private information or identifiable biospecimens that might be used in research, whether sharing of the information or biospecimens might occur, and the types of institutions or researchers that might conduct research with the information or biospecimens</td>
</tr>
<tr>
<td>16. A statement describing how long the information or biospecimens may be stored and maintained and how long the information or biospecimens may be used for research purposes; these time periods may be indefinite</td>
</tr>
<tr>
<td>17. A statement that clinically relevant research results may not be disclosed to the participant</td>
</tr>
<tr>
<td>18. A statement informing the participant whom to contact for answers to questions about the participant’s rights regarding storage and use of information or biospecimens and whom to contact regarding research-related harm</td>
</tr>
</tbody>
</table>

Traditionally, in-person paper consents are often resource intensive, not easily scalable, and preclude digital responses from being incorporated in the electronic health record and laboratory information management systems. Given that PM requires large-scale patient engagement, innovations in consenting in conjunction with broad public education [5,6] are required. The emergence of digital health plays a substantial role in defining population-based approaches to electronic consent. Interactive and multimedia slideshow consents have been used for enrollment of participants in biobanks [7,8], but such slideshow consents require increased participant time. Animated video consent approaches have been effective in
providing comprehensive information and improving participants’ understanding of content [9,10], but such video consents have not been used in biobank research associated with PM.

EngageUC, an NIH-funded study, examined biobanking in the University of California system with community constituents to better define the innovative and accessible consent materials needed as part of a scalable institutional biobanking program in support of PM [4,11]. The following key themes emerged: the public should be educated about biobanking, consent content source should be considered knowledgeable and trustworthy, consent process should be low stress with an opportunity to get answers to questions, format and language of the consenting material should be easy to understand, and oversight should be conducted by the community and stakeholders.

In this study, we engaged the community and stakeholders across the University of California, Los Angeles (UCLA) Health System, David Geffen School of Medicine at UCLA, UCLA Institute of Precision Health, and the UCLA Clinical and Translational Science Institute (CTSI) to create and pilot an innovative potentially scalable universal video consent that asks patients to give a “broad” or “one time” consent that allows researchers to use their biomaterial and clinical data in a manner that meets the criteria defined by both the NIH and NPRM [1].

Methods

This study was approved by the UCLA Institutional Review Board (IRB; #15-001395IRB) with a waiver of written informed consent.

Governance Structure

We formulated a strong governance structure, including a community advisory board (CAB), to oversee and give feedback on the consent design and process [11].

Community Advisory Board

Our study team assembled a CAB consisting of 11 respected leaders that were highly involved with organizations in the Los Angeles region that understood our diverse communities and represented their perspectives. The members were racially diverse (2 were African American, 2 were Latinx, 1 was Asian American, 1 was Native American, 1 was Persian-American, and 4 were non-Hispanic White) and equitable with respect to gender (5 were males and 6 were females).

The committee held five meetings between July 2015 and June 2019 to review and guide the video design. The CAB’s primary focus was to ensure the video was easy to understand and explained the purpose of the consent. The board additionally focused on three key areas: inclusion of diverse patients in PM education, outreach, and research; integration of research and clinical operations; and potential return of genetic results.

Internal Advisory Board

The internal advisory board included our institutional research leaders from the David Geffen School of Medicine, Institute of Precision Health, CTSI, UCLA Health, and additional faculty with expertise in bioethics, patient engagement, biobanking, and IRB. The members of the internal advisory board provided substantial feedback to ensure the video content was informative, met NIH standards, addressed both genetic testing and the potential for collaborations with external companies and federal partners, was culturally sensitive, and represented the diversity of Los Angeles.

Video Development

The content for both the text and animated video consents were adapted from paper versions of a biobanking consent developed by EngageUC [4]. The animated consent included a statement about collaboration with governmental agencies, commercial entities, and other academic institutions, and a statement that potential secondary use of data could include genomic sequencing. The videos were targeting a seventh grade reading level. Both video consents were designed to be 4 to 5 minutes in length. These pilot videos were in English and Spanish, with voiceovers for the animated consent. All the essential components for an NIH informed consent were included in the videos (see Textbox 1) [12].

Phase I

A text-based video (text moving from screen to screen) was first designed to consent patients around the use of their remnant tissue for research. A convenience sample of 125 patients were enrolled but only 123 completed postconsent surveys (see Figure 1).

Figure 1. Pilot-testing electronic universal consent (EUC).
**Phase 2**

Our CAB and internal advisory board guided the adaptation of the universal consent video to a fully animated (cartoon-like) video to better communicate content to lay audiences and use this to power the Institute of Precision Health ATLAS biobank (see Figure 2). The video conveyed that this sample would be collected at one time and as a piggyback to any standard lab draw or intravenous placement. For this phase, an additional convenience sample of 50 patients were enrolled, of whom 47 completed postconsent surveys (see Figure 1). Phase II pilot testing was mainly to evaluate if the additional animation improved the user experience of the consent video.

**Consenting Process**

The first pilot was conducted at five distinct locations within UCLA Ronald Reagan Hospital: (1) hospital admissions, (2) clinical lab, (3) mammography clinic, (4) oncology clinic, and (5) liver clinic. Technical assistance was available at all these locations. Sites were selected because of their diverse populations and high volume (e.g., mammography). All patients were approached and technical assistance from study staff was made available. The second pilot was expanded to include perioperative suites. Patients had to validate their identity before viewing the video and responding to the consent questions (see Figure 2).

Patients were asked to choose their preferred language (English or Spanish) and then validate their identity by entering their medical record number, selecting their birth year (out of 6), and entering the initials of their first and last name. Once validated, individuals could view the video and provide consent. There was no prompting from study staff or clinic personnel. A paper brochure with frequently asked questions (FAQs; available in English and Spanish) was handed to patients with the iPad. Both the video and the FAQs let patients know they could change their consent status at any time. After watching the video, patients were asked: where they wanted to opt in or opt out of having remnant biospecimens used for research and if they would be open to recontact for future research.

**Data Collection and Outcome Measures**

Demographic data, including age, race or ethnicity, highest level of education, and language preference for the convenience sample, were collected.

A postconsent survey was developed in English and Spanish to evaluate the effectiveness of the universal consent videos and understand drivers of consent choice. Patients were approached after completing the universal consent videos. Volunteers received a US $5 dollar Target gift card to compensate them for their time. *Patient impressions* of the universal consent videos were evaluated with three questions (how useful did you find the information, how easy was the information to understand, and how much did you feel you could trust the information) using a five-point Likert scale (not at all, not really, somewhat, mostly, and very).

**Drivers of Consent Choice**

The internal advisory board helped develop questions used to understand the reasons patients’ opted in or out. Individuals who opted in received additional questions to determine drivers (hoping the research will help me in the future, hoping the research will help my family and friends in the future, hoping to advance science, or hoping to find a cure for a disease). Individuals who opted out received additional questions to determine drivers (do not want my tissue used for anyone else, concerns about privacy, concerns that a product will be made...
with my tissue and I will not benefit, or did not understand what I was being asked).

**Patient Health Beliefs Regarding Medical Research**

The study team developed a 10-item questionnaire to evaluate patients’ beliefs about biomedical research using validated instruments as guides [13-16]. This new survey measured attitudes about science, optimism, altruism, privacy, social support, justice, and conflict of interest.

**Opt-in or Opt-out Status**

The patient’s decision to either opt in or opt out of sharing their remnant samples was recorded. We also tracked the number of patients who agreed to be contacted for future research.

**Statistical Analyses**

Demographic information, consent rates, and patient ratings about consent were summarized using descriptive statistics including medians, ranges, and percentages. The video script was run through the Flesh-Kinkaid Readability Test tool in Word (Microsoft Corporation) to determine the grade level of the universal consent videos. The patient characteristics between those that opted in and opted out were compared with chi-square tests for homogeneity. The consent rates were compared between phases 1 and 2 consents with chi-square tests for homogeneity. The consent rates were compared between phases 1 and 2 with chi-square tests for homogeneity. Patients’ evaluation of the usefulness, ease, and trustworthiness of the videos in phases 1 and 2 were compared using Wilcoxon rank sum tests since the variables had skewed distributions. To determine the internal consistency of the 10-item beliefs survey, Cronbach alpha was used. Univariate logistic regression was used to examine the association between patients’ beliefs and their consent decision. We used the AUC receiver operating characteristic curves to predict which patients opted in versus opted out. Two-sided $P$ values were reported, and variables were considered statistically significant if the $P$ value was <.05. All analyses were conducted using Stata 15 (StataCorp) [17].

**Results**

**Community Advisory Board Suggestions**

The CAB played a significant role in the video design, which made the video rich in content, ensured that the language was appropriate for the lay population, addressed the most concerning questions from the community, was applicable to a diverse population, and was less than 5 minutes in duration. The board’s primary focus was to ensure that the video was understandable and appropriately explained why UCLA was asking them to donate biosamples and clinical data for research. The board additionally focused on three key areas: (1) how to ensure the inclusion of diverse patients and communities in PM program education, outreach, and research; (2) if and how to return PM research findings to individual patients who contribute samples and data to the biobank; and (3) how to appropriately bridge research and clinical operations.

**Participants**

A total of 175 patients enrolled across the two pilot phases, of which 173 actually went through the electronic video consent (see Figure 1). The population was mostly middle age (n=130, 75% were younger than 60 years), female (n=123, 69%), White (n=86, 50%), and educated (n=104, 60% had at least college education; Table 1). The majority of patients preferred English (n=161, 93%). There were no significant differences for age, education, gender, or race between patients who opted in or opted out.
Table 1. Sociodemographic data (cohort that tested consent).

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Participants (N=173), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;30</td>
<td>36 (20.8)</td>
</tr>
<tr>
<td>30-39</td>
<td>37 (21.4)</td>
</tr>
<tr>
<td>40-49</td>
<td>25 (14.5)</td>
</tr>
<tr>
<td>50-59</td>
<td>33 (19.1)</td>
</tr>
<tr>
<td>60-69</td>
<td>20 (11.6)</td>
</tr>
<tr>
<td>70-79</td>
<td>20 (11.6)</td>
</tr>
<tr>
<td>≥80</td>
<td>2 (1.2)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>50 (30.6)</td>
</tr>
<tr>
<td>Female</td>
<td>123 (69.4)</td>
</tr>
<tr>
<td><strong>Race/ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>86 (49.7)</td>
</tr>
<tr>
<td>Asian</td>
<td>29 (16.9)</td>
</tr>
<tr>
<td>Black</td>
<td>21 (12.2)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>31 (18.0)</td>
</tr>
<tr>
<td>Native American</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Other</td>
<td>5 (2.3)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>Less than high school</td>
<td>8 (4.7)</td>
</tr>
<tr>
<td>High school graduate</td>
<td>30 (17.7)</td>
</tr>
<tr>
<td>Some college</td>
<td>28 (16.5)</td>
</tr>
<tr>
<td>College graduate</td>
<td>53 (31.2)</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>27 (15.9)</td>
</tr>
<tr>
<td>MD or PhD</td>
<td>24 (14.1)</td>
</tr>
<tr>
<td>Unknown</td>
<td>3 (1.7)</td>
</tr>
<tr>
<td><strong>Language</strong></td>
<td></td>
</tr>
<tr>
<td>English preferred</td>
<td>161 (93.1)</td>
</tr>
<tr>
<td>Spanish preferred</td>
<td>12 (6.9)</td>
</tr>
</tbody>
</table>

**Consent Rate**
There was no significant difference for consent rate between the two phases (44/50, 88.0% vs 112/123, 91.1%; \( P = .41 \)). Across the entire cohort, 56% (97/173) of individuals agreed to be recontacted to participate in other biomedical research projects.

**Patients’ Health Beliefs on PM Research**
The 10-item questionnaire had good internal consistency with an alpha coefficient of .93, which means the results were consistent among similar questions. Univariate logistic regression analysis showed that there were significant differences on all 10 items between the groups who opted in versus opted out (all \( P < .001 \)). Additionally, we calculated AUC to evaluate the ability of the questions to discriminate which question predicted patients opting in. The results showed that “I trust researchers to use leftover biological specimens to promote the public’s health” and “Sharing a biological sample for research is safe because of the privacy protections in place” discriminated opt-in statuses were the strongest predictors (both AUC were 0.88; Table 2).
Table 2. The association between participants’ health beliefs and demographic characteristics with their decision to opt in (N=164).

<table>
<thead>
<tr>
<th>Health beliefs and demographics</th>
<th>Construct</th>
<th>Odds ratio (95% CI)</th>
<th>AUC&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. Results of research using biological samples will help future generations.</td>
<td>Attitude toward science, optimism, altruism</td>
<td>5.5 (2.6-12.2)*</td>
<td>0.8*</td>
</tr>
<tr>
<td>Q2. It is important for individuals to participate in research to advance science.</td>
<td>Altruism, communitarianism</td>
<td>10 (4.1-24.5)*</td>
<td>0.87*</td>
</tr>
<tr>
<td>Q3. Sharing a biological sample for research is safe because of the privacy protections in place.</td>
<td>Attitude toward science, privacy concerns</td>
<td>4.5 (2.4-7.5)*</td>
<td>0.88*</td>
</tr>
<tr>
<td>Q4. Results of the research using donated biological samples will help me or my family in the future.</td>
<td>Attitude toward science, optimism, altruism</td>
<td>4.5 (2.3-8.3)*</td>
<td>0.84*</td>
</tr>
<tr>
<td>Q5. My family and friends support donating biological samples for research.</td>
<td>Social support</td>
<td>3 (1.7-5.4)*</td>
<td>0.8*</td>
</tr>
<tr>
<td>Q6. Research on donated tissue may lead to medical breakthroughs from which UCLA&lt;sup&gt;b&lt;/sup&gt; and researchers will profit.</td>
<td>Attitude toward science, justice, conflict of interest</td>
<td>3 (1.7-5.1)*</td>
<td>0.76*</td>
</tr>
<tr>
<td>Q7. I trust researchers to use leftover biological specimens to promote the public’s health.</td>
<td>Attitude toward science, justice, trust</td>
<td>5.5 (2.7-11)*</td>
<td>0.88*</td>
</tr>
<tr>
<td>Q8. The most important thing to researchers is helping people and curing disease.</td>
<td>Attitude toward science</td>
<td>5 (2.4-10)*</td>
<td>0.79*</td>
</tr>
<tr>
<td>Q9. People have a responsibility to help each other.</td>
<td>Altruism</td>
<td>5.5 (2.5-11)*</td>
<td>0.8*</td>
</tr>
<tr>
<td>Q10. If a person does not donate tissue for research it just goes to waste.</td>
<td>Attitude toward science</td>
<td>3.3 (1.9-6.2)*</td>
<td>0.81*</td>
</tr>
<tr>
<td>Age</td>
<td>N/A&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1.3 (0.9-1.7)</td>
<td>0.62</td>
</tr>
<tr>
<td>Education</td>
<td>N/A&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.96 (0.7-1.3)</td>
<td>0.52</td>
</tr>
<tr>
<td>Race</td>
<td>N/A&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td></td>
<td>1.00 (reference)</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td></td>
<td>0.56 (0.17-1.84)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td></td>
<td>0.49 (0.13-1.78)</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td></td>
<td>1.76 (0.36-8.66)</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td>0.36 (0.03-3.89)</td>
<td></td>
</tr>
<tr>
<td>Gender (female)</td>
<td></td>
<td>0.7 (0.3-1.9)</td>
<td>0.54</td>
</tr>
</tbody>
</table>

<sup>a</sup>AUC: area under the curve.
<sup>b</sup>UCLA: University of California, Los Angeles.
<sup>c</sup>N/A: not applicable.
<sup>*</sup>P<.001

Evaluation of the Universal Consent Video

We also examined whether there was a difference between the video consent evaluations of patients who opted in and opted out regarding the ease of use, usefulness, and the trustworthiness as three outcomes: useful and not useful, easy to understand and not easy to understand, and trustworthy and not trustworthy, respectively. In terms of where it was useful or easy to understand, the universal consent video did not differ between two groups (those who opted in vs opted out). However, 88.4% (136/158) of the patients who opted in felt they could trust the information in the video compared to only 53.3% (8/15) of the patients who opted out (P<.001).

We compared the evaluations of the phase 1 text-based video and the phase 2 animated video among patients regarding the ease of use, usefulness, and trustworthiness. We found that there was a statistically significant difference between the text-based video and the animated video regarding the ease of use (P=.005) and the ability to understand this information (P<.001). There was no significant difference regarding the trustworthiness between the text-based video and animated video (P=.20; Table 3).

https://formative.jmir.org/2021/9/e29123

JMIR Form Res 2021 | vol. 5 | iss. 9 | e29123 | p.138
(page number not for citation purposes)
Table 3. Comparison of usefulness, ease of use, and trustworthiness between two pilot phases of video consent (N=164).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Phase 1 (n=117)</th>
<th>Phase 2 (n=47)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usefulness</td>
<td>4 (3-5)</td>
<td>5 (4-5)</td>
<td>.005</td>
</tr>
<tr>
<td>Ease</td>
<td>5 (4-5)</td>
<td>5 (5-5)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Trustworthiness</td>
<td>4.5 (4-5)</td>
<td>5 (4-5)</td>
<td>.20</td>
</tr>
</tbody>
</table>

Responses were based on a 5-point Likert scale: 1 (not at all), 2 (not really), 3 (somewhat), 4 (mostly), and 5 (very).

Important Factors for Opting In and Opting Out

Questions that garnered a large majority of patients (≥80%) responding as “moderate” or “very important” were a key focus. Among the four questions we asked the patients who opted in (Table 4), three of four made this threshold: “research benefiting me,” “hoping PM research could advance science,” and “cure diseases.” Among the four questions we asked the patients who opted out (Table 5), only 1 question about “privacy” made this threshold and was a factor for all the patients.

Table 4. Reasons for opting in (n=101 completed).

<table>
<thead>
<tr>
<th>Reasons for opting in</th>
<th>Not at all, n (%)</th>
<th>A little, n (%)</th>
<th>Moderate important, n (%)</th>
<th>Very important, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoping the research will help me in the future</td>
<td>6 (5.9)</td>
<td>10 (9.9)</td>
<td>27 (26.7)</td>
<td>58 (57.4)</td>
</tr>
<tr>
<td>Hoping the research will help family, friends, or others</td>
<td>0 (0)</td>
<td>4 (4)</td>
<td>13 (12.9)</td>
<td>84 (83.1)</td>
</tr>
<tr>
<td>in the future</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoping to advance science</td>
<td>0 (0)</td>
<td>2 (1.9)</td>
<td>9 (8.9)</td>
<td>90 (89.1)</td>
</tr>
<tr>
<td>Hoping to contribute to the cure of diseasež</td>
<td>0 (0)</td>
<td>2 (2)</td>
<td>4 (4)</td>
<td>94 (94)</td>
</tr>
</tbody>
</table>

One patient did not answer the question.

Table 5. Reasons for opting out (n=20 completed).

<table>
<thead>
<tr>
<th>Reasons for opting out</th>
<th>Not at all, n (%)</th>
<th>A little, n (%)</th>
<th>Moderate important, n (%)</th>
<th>Very important, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do not want my tissue used for anyone elsež</td>
<td>6 (31.6)</td>
<td>2 (10.5)</td>
<td>6 (31.6)</td>
<td>5 (26.3)</td>
</tr>
<tr>
<td>Concern about privacy</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>2 (10)</td>
<td>18 (90)</td>
</tr>
<tr>
<td>Concern that a product may be made from my tissue and I</td>
<td>5 (26.3)</td>
<td>4 (21.1)</td>
<td>2 (10.5)</td>
<td>8 (42.1)</td>
</tr>
<tr>
<td>will not benefitž</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did not understand what I was asked to consent tož</td>
<td>8 (44.4)</td>
<td>4 (22.2)</td>
<td>4 (22.2)</td>
<td>2 (11.1)</td>
</tr>
</tbody>
</table>

One patient did not answer the question.

Two patients did not answer the question.

Discussion

Our study indicated that our universal consent animated video is easy and informative. Because it is short and self-administered, this is a possible solution for a scalable consent method for population-based PM research. Compared to in-person paper consent, electronic video consent requires fewer human resources and less physical space. As designed in this study, it could be deployed to any number of devices and applied at multiple medical locations. Hence, it is suitable for large-scale efforts to collect informed consent from a large population with a modest incremental cost. Furthermore, it allows patients a safe space to participate in the consenting process without the pressure an in-person process might create. To apply it broadly and effectively to diverse populations, it is critical that the universal consent video addresses potential concerns participants may have about the research project to build trust, reassure potential participants about privacy concerns, be transparent (which further increases trust), and address the potential of the research.

In line with other studies, we found that trust is one of the most important factors for patients opting in to biomedical research [18]. Multiple studies have identified reasons for reduced trust between patients and researchers: participants are not clear about their rights over their data in the biobank [19]; patients did not understand biobanking or the aims of the clinical trial [20]; patients might have concerns about allowing researchers to use their data for the unforeseen secondary research via a broad consent process [21]; patients who consented to participate in clinical trials heavily depended on how much they trusted the physician [22], whereas in this consent process, there are no health professionals communicating with patients; or there is no immediate benefit for patients in PM research.

Delivering comprehensive information about biobanking and PM research is necessary for truly informed consent and to build patients’ trust. However, it is important to balance the video content and length, as patients might lose interest or read or watch the consent cursorily if it is too long or if the content is not presented in language that average individuals can understand [23-25]. One solution to increase patients’
understanding of and trust in PM research and biobanking may be to provide more concrete examples of clinical research and PM. A complementary approach may be to provide personal stories of successful PM in UCLA patients. Such educational videos could help interested individuals learn more about the value of remnant biospecimens, clinical data, and clinical research in advancing science. It is important to ensure patients understand that PM research takes time, so the benefits of participation will not be immediate.

In this study, all (100%) patients who opted out responded that concerns about privacy were moderately or very important to them. This is consistent with results from multiple studies suggesting that patients were concerned about misuse of their personal data [26]. If patients do not understand how their data might be used or who might use the data, they are less likely to give permission to share the data [27]. As PM research requires hundreds and thousands (or more, depending on the specific question) of unique biological samples, emphasizing how clinical information will be protected should be embedded in the consent process. Furthermore, a transparent policy to efficiently manage data access and protect individual’s privacy through a variety of data access controls and an oversight committee for ethical governance of the biobank is a necessity [4]. In fact, some authors believe this represents the only way to build public trust and protect participants’ privacy [28]. Researchers, scientists, and policy makers should embrace the notion that if privacy concerns are well addressed in the consent and clearly communicated in a trustworthy way, this could enhance potential participants’ understanding of and trust in the research process.

Our study found that potential participants’ health beliefs were the most significant driver of their willingness to participate in a precision health initiative. Patients who opted in believed that their participation could advance science, find cures for disease, and help others. This confirms previous studies that participation in biobank research was based on altruistic motivations and responsibilities to assist future generations [18]. Together with early studies, our findings suggested that emphasizing the importance of patients’ participation to benefit others and contribute to science is associated with the high participation rate in clinical research. From these 10 health belief questions, we again confirmed that if patients trust the researchers and believe their personal privacy is protected, they are more likely to donate their biospecimens.

This pilot study has its limitations. This study only included a convenience sample of patients who agreed to do the electronic consent and answer the additional survey. The sample size was small, and there were smaller subgroups in each category of race, age, gender, and educational level, which limited our ability to evaluate any differences between these populations. We also did not evaluate the participants’ health statuses, which prevented us from understanding if differences in consent rate and health beliefs exist among patients with different diagnoses or disease burden. Future research needs to evaluate the electronic video consent performance in a larger population, so these and other potentially important variables such as low health literacy can be studied more comprehensively.

In summary, we created and piloted an innovative electronic video consent that was self-administered and easy to understand for patients. This approach will next be tested for scalability as an enterprise solution by expanding across 18 clinical sites across the UCLA health system. Future goals include expansion to other University of California sites and piloting the video and process in affiliated county hospitals within the larger Los Angeles County. We believe our video consent and process offer an approach that would allow for more robust inclusion of institutions that do not have the financial resources to use employees for in-person consent. Given the reality that many such institutions will serve patients who are chronically ill, of lower socioeconomic status, and who are from underrepresented minority populations, our video consent and process offer the possibility for these groups to become better represented in PM research. The importance of participation in PM remain unclear especially among ethnic minority populations.

Acknowledgments

The Clinical Translational Science Institute and Institute for Precision Health at UCLA wish to thank all the patients who participated in the study. The authors would also like to acknowledge all the members in the CAB.

The study was supported by Federal Fund, NIH-NCATS National Center for Advancing Translational Science (award number UL1TR001881-01S2; grant title: UCLA Clinical Translational Science Institute).

Conflicts of Interest

None declared.

References


17. Stata. College Station, TX: StataCorp LP; 2015. URL: https://www.stata.com/ [accessed 2021-06-06]


Abbreviations

- AUC: area under the curve
- CAB: community advisory board
- CTSI: Clinical and Translational Science Institute
- FAQ: frequently asked question
- IRB: Institutional Review Board
- NIH: National Institutes of Health
- NPRM: Notice of Proposed Rule Making
- PM: precision medicine
- UCLA: University of California, Los Angeles

©Arash Naeim, Sarah Dry, David Elashoff, Zhuoer Xie, Antonia Petreuse, Clara Magyar, Lilliana Johansen, Gabriela Werre, Clara Lajonchere, Neil Wenger. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.09.2021. 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.
Original Paper

Social Networking Site Use During the COVID-19 Pandemic and Its Associations With Social and Emotional Well-being in College Students: Survey Study

Alison B Tuck¹, MA; Renee J Thompson¹, PhD
Department of Psychological and Brain Sciences, Washington University in St Louis, St Louis, MO, United States

Corresponding Author:
Alison B Tuck, MA
Department of Psychological and Brain Sciences
Washington University in St Louis
Campus Box 1125
One Brookings Drive
St Louis, MO, 63130
United States
Phone: 1 (314) 935 3502
Email: alison.tuck@wustl.edu

Abstract

Background: Social distancing during the COVID-19 pandemic has reduced the frequency of in-person social interactions. College students were highly impacted, since many universities transferred curriculum from in-person to entirely online formats, physically separating students with little notice. With social distancing, their use of social networking sites (SNSs) likely changed during the COVID-19 pandemic, possibly holding implications for well-being.

Objective: This study aimed to determine (1) how components of SNS use (ie, weekly frequency, time per day, habitual use, engagement, enjoyment, addiction, and emotional impact) changed from before to during COVID-19, (2) how these changes in SNS use were associated with pandemic-related social and emotional well-being, and (3) how SNS use and changes in use during the pandemic were associated with loneliness.

Methods: College students (N=176) were surveyed during the time when their university campus in the United States was operating online. Participants completed the same SNS use questionnaires twice, once with regard to the month preceding the onset of COVID-19 and again with regard to the month since this time. They also reported the extent to which they experienced perceived change in social support resulting from the pandemic, pandemic-related stress, and general loneliness.

Results: After the onset of COVID-19, participants showed an increase in daily time spent on SNSs (t₁₆₉=5.53, d=0.42, P<.001), habitual use (t₁₇₃=3.60, d=0.27, P<.001), and addiction (t₁₇₃=4.96, d=0.38, P<.001); further, enjoyment on SNSs decreased (t₁₇₃=−2.10, d=−0.16, P=.04) and the emotional impact of SNS activities became more negative (t₁₇₂=−3.76, d=−0.29, P<.001). Increased perceived social support during COVID-19 was associated with changes in frequency of SNS use, time per day, addiction, and engagement (r>0.18 for all). Pandemic-related stress was associated with changes in SNS addiction and the extent to which one’s SNS content was related to the pandemic (r>0.20 for all). Loneliness was positively associated with SNS addiction (r=0.26) and negatively associated with SNS engagement (r=−0.19) during the pandemic. Loneliness was also negatively associated with changes in habit and engagement (r=−0.15 for all).

Conclusions: Findings suggest that components of SNS use are associated with both positive and negative pandemic-related social outcomes, but largely negative pandemic-related emotional outcomes. Further, some components of SNS use are positively associated with loneliness (eg, addiction) while others show a negative association (eg, engagement). These findings provide a more nuanced picture of how SNS use is associated with social and emotional well-being during the time of a global health crisis when in-person interactions are scarce.

(JMIR Form Res 2021;5(9):e26513) doi:10.2196/26513

KEYWORDS

social media; social networking sites; COVID-19; loneliness; well-being
Introduction

The infectious respiratory disease COVID-19 was first recognized in Wuhan, China, in December of 2019 [1]. On March 11, 2020, the World Health Organization declared COVID-19 a pandemic. As of February 17, 2021, COVID-19 has infected over 109 million individuals across the globe, and over 2.4 million deaths caused by COVID-19 have been reported to date [2]. Approximately a quarter of these cases are from the United States, representing the largest number of cases compared to any other country across the globe [2].

Like many widespread outbreaks of infectious diseases [3], the COVID-19 pandemic has been associated with negative psychological outcomes, such as increased rates of depression, anxiety, and stress in the general population [4]. Several researchers theorize that these mental health outcomes are a result of social distancing—the act of physically isolating and not interacting in person with others to reduce the risk of spreading the disease [5,6]. A possible explanation of this link is that less frequent in-person social interactions are associated with lower psychological well-being [7]. With social distancing being recommended and sometimes enforced in communities across the globe, it is important to examine how individuals are coping with, and may be compensating for, their less frequent in-person interactions. In a technology-driven world, social networking sites (SNSs) may be the best alternative for many people. The central goals of this investigation are to examine how SNS use has changed during the COVID-19 pandemic and to analyze how these changes in SNS use are related to pandemic-related social and emotional well-being as well as loneliness.

SNSs refer to a specific type of social media in which “communities” are formed consisting of public or semipublic profiles and where individuals can regulate with whom they connect as well as browse the connections of others [8]. Use of SNSs has been assessed with a variety of measures, most of which almost exclusively assess time per day spent on SNSs. Studies using these measures have yielded mixed results [9]. Some studies show positive associations between time spent on SNSs and negative mental health outcomes (ie, depression and anxiety) [9,10]. Other studies show no such associations. For example, a recent 8-year longitudinal study found no association between daily time spent on SNSs and depression or anxiety [11], stressing the need for researchers to evaluate use of SNSs beyond a focus on screen time.

In addition to examining time spent on SNSs, Turel and Serenko [12] developed a model for understanding a wide range of components of SNS use and how they lead to either favorable or adverse outcomes: SNS addiction, engagement with SNSs, time per day, enjoyment on SNSs, and habitual SNS use. According to their model, SNS addiction—defined as a dependency on SNSs that results in an obsessive pattern of SNS seeking and use that interferes with engagement in other important activities—is an adverse and pathological outcome [12]. Conversely, engagement with SNSs—defined as individuals caring about SNSs and making them a significant aspect of their lives that they can control—is seen as a favorable, nonpathological outcome [12].

Turel and Serenko theorize that enjoyment on SNSs, or an individual’s intrinsic motivation for using SNSs simply because of their emotional rewards, is what leads to high SNS engagement [12]. However, enjoyment on SNSs seems to be a double-edged sword, as this variable, along with time per day spent on SNSs, is theorized to lead to habitual SNS use. Turel and Serenko hypothesize that habitual use occurs when individuals use SNSs automatically in certain contexts due to some learned association. This habitual use can often lead to SNS addiction, the pathological outcome. Consistent with their theorizing, SNS addiction has been found to be positively associated with decreased psychological well-being (eg, depressive symptomology) [13]. In this regard, Turel and Serenko’s model provides a detailed overview of the use and consequences of SNSs. However, researchers have not analyzed how these various SNS use components may have changed as a result of social distancing caused by widespread disease. With current social distancing initiatives making it so that SNS use may be one of the most common forms of social interaction, individuals might be using SNSs more during the pandemic while possibly not reaping the same emotional benefits from its use.

An important next step for SNS research is to examine how SNS use is associated with emotional experiences. Analyzing the emotions experienced by individuals on SNSs expands the literature by clarifying when in-the-moment SNS use might be positive and when it might be negative. This moves the field beyond measuring associations between SNS use and depression and anxiety—symptoms of disorders that have relatively low base rates—and allows us to analyze the short-term emotional influence of SNSs and how SNSs affect quality of life. For example, perhaps there are periods of time in which individuals experience more positive emotions while on SNSs, and other times in which they experience more negative emotions. These short-term emotional impacts, when experienced regularly, could have important implications for psychological well-being.

Emerging literature demonstrates that SNS use is associated with negative emotional experiences during the current pandemic. For instance, weekly frequency of exposure to COVID-19–related content on SNSs was associated with higher levels of general psychological distress in a large Chinese sample [14]. Some researchers have postulated that the misinformation being spread across SNSs about the pandemic (eg, conspiracy theories) leads to increases in stress, anxiety, and panic in relation to COVID-19 [14-16]. Evidence and theory that exposure to COVID-19–related SNS content correlates with negative mental health outcomes suggests that SNS use during COVID-19 might be associated with decreases in social and emotional well-being during the pandemic.

Another possible negative consequence of the pandemic is increased rates of loneliness. Indeed, a recent large-scale study of adults found that approximately 36% of participants endorsed sometimes or often feeling lonely during the pandemic [17]. Loneliness is conceptualized as a wish to feel closer to others when individuals otherwise feel isolated [18]. This sense of
isolation may be the result of being physically separate from others—as could particularly be the case during social distancing initiatives during COVID-19—or of feeling emotionally isolated. Of note, compared to midlife and older adults, younger adults are particularly vulnerable to experiencing loneliness when they have a diminished quantity (versus quality) of social engagements [19]. Given that many college students experienced their universities abruptly cease in-person operations during COVID-19, this group experienced great decreases in social interactions and likely experienced increases in loneliness.

Ample research suggests that loneliness is associated with SNS use. For example, those who have few in-person social interactions and who use SNSs a great amount report higher levels of loneliness than other groups of individuals, including those who have few in-person social interactions and who use SNSs only a small amount [20]. Research testing causal models suggest that loneliness is the cause of increased SNS use and not that SNS use is the cause of loneliness [21]. Taken together, this literature suggests that loneliness may be associated with particularly high increases in college students’ SNS use during the pandemic, when in-person social interactions are scarce.

This investigation had three primary aims. The first aim was to determine how SNS use changed, overall, from before to after the onset of the COVID-19 pandemic. Consistent with Turel and Serenko’s model [12], we hypothesized that components of SNS use associated with addiction—which is associated with worse psychological well-being—would increase from pre- to during COVID-19. Specifically, compared to pre-COVID-19, we hypothesized that time per day spent on SNSs, habit, and addiction increased during COVID-19. In addition, we expected that SNS engagement decreased, and we did not expect enjoyment to significantly change during this period. Consistent with prior research and theory, we also expected that SNS frequency increased and that the average emotional impact from SNS use decreased during COVID-19.

The second aim was to examine how changes in SNS use were associated with pandemic-related social well-being (ie, changes in perceived social support during the pandemic) and pandemic-related emotional well-being (ie, pandemic-related stress). Again, consistent with Turel and Serenko’s model and other research, we expected that poor pandemic-related social and emotional well-being would be positively associated with changes in SNS frequency, time per day, habit, addiction, and the percentage that one’s SNS content was related to the pandemic. Further, we hypothesized that lower levels of pandemic-related social and emotional well-being would be associated with decreases in SNS engagement and average emotional impact from SNSs. Importantly, we expected each of the associations between SNS use and pandemic-related emotional well-being to remain significant after controlling for general distress, which was indexed by depressive symptoms.

To expand upon the extant literature regarding associations between SNS use, COVID-19, and loneliness, the third aim was to examine the associations between loneliness and both (1) components of SNS use during COVID-19 and (2) changes in components of SNS use from pre- to during COVID-19. In this investigation, loneliness was conceptualized as an outcome that was not specifically related to the pandemic. In other words, although we expect that loneliness increased during the pandemic, loneliness was examined as a general measure. Given that loneliness is also a type of psychological distress indicative of poor psychological well-being, consistent with Turel and Serenko’s model, we hypothesized that loneliness would be positively associated with changes in time spent on SNSs, frequency of SNS use, habitual SNS use, and SNS addiction and that it would be negatively associated with engagement with SNSs, enjoyment on SNSs, and the average emotional impact of SNSs. We expected parallel associations of SNS use during COVID-19 specifically (eg, loneliness would be positively associated with time spent on SNSs during COVID-19). Importantly, we expected that all of the relationships between loneliness and SNS use components would hold even after accounting for social anxiety, which is positively associated with loneliness [22].

This study focused on SNS use among college students. Approximately 90% of young adults in the United States aged 18 to 29 years use SNSs, representing the largest adult group to engage with these platforms [23]. Further, college students shared unique experiences early during the COVID-19 outbreak in that colleges closed midterms across the United States. As a result, college students were specifically and greatly impacted by disturbances to their normal social functioning during the pandemic, possibly above and beyond any other adult group. We capitalized on this clearly defined disruption among this group (ie, before universities closed versus during university closures) to examine the impact of COVID-19 on SNS use. Finally, college students are at an increased risk for various psychological disorders, including depression, anxiety, and substance use disorders, at rates higher than their older peers [24,25]. In addition, significantly more college-aged adults endorse serious psychological distress, such as feeling nervous or hopeless, compared to adults aged 22 to 34 years [26]. This trend may, in part, be explained by college students facing many unique stressors, such as academic pressure and first-time separation from family [26,27]. Of note, although college students are not at increased risk for developing a psychiatric disorder compared to their peers who do not attend college, they are significantly less likely to receive mental health treatment [28,29]. The shortage of mental health treatment available to college students has been deemed to be a mental health crisis [30,31]. SNSs may provide a needed venue for college students to engage in self-disclosure and establish social connection when they cannot acquire formal mental health treatment [32]. For these reasons, we think examining college students during the COVID-19 pandemic will provide a more thorough understanding of the role of SNSs during this challenging time.

Methods

Recruitment

The entire study was administered online from April 14 to 24, 2020. Undergraduate students in psychology courses learned about the study via a university portal that lists active studies. The portal was open to all undergraduate students enrolled in psychology courses at the university, and it provided students...
with a hyperlink to access the study. The first webpage of the study presented interested individuals with an informed consent form. Those who consented were directed to complete a demographics questionnaire followed by the rest of the study measures. All participants completed the study within a time frame of about one hour and received one hour of course research credit for their participation. All study procedures were reviewed and approved by the Institutional Review Board at Washington University in St Louis, Missouri.

**Procedures**

Cases of COVID-19 surged in the United States in March of 2020. Coincidentally, this initial surge began during the university’s spring break, which took place from March 8 to 21, 2020, when almost all students leave campus. During this time, students were told that the university would no longer be holding in-person instruction, and they were not allowed to return to campus. As such, all participants in this study shared the same unique experience of not just being students who use SNSs quite regularly, but also of COVID-19 having the largest impact on daily life after spring break with a clear date delineating “pre–COVID-19” and “during COVID-19” time frames.

In this one-part study, we administered four sets of measures. First, we administered the same series of SNS use self-report measures twice; the only difference was the period of time that students considered when completing them. For the first set of SNS use measures, students answered with regard to the month preceding their spring break (ie, “pre–COVID-19,” from February 7 to March 7, 2020, before receiving the news that instruction was transitioning online). For the second set of SNS use measures, they answered with regard to the time since spring break (ie, “during COVID-19”), which ranged anywhere from 3 weeks and 1 day earlier to 4 weeks and 4 days earlier. In this second set of SNS use questions, participants were additionally asked to report on the extent to which the content on their SNSs was related to the pandemic. The third set of measures included three measures assessing pandemic-related social and emotional well-being. Finally, the fourth set of measures included three psychological distress measures, the order of which were randomly presented across participants.

**Measures**

**Components of Social Networking Site Use**

**Overview**

We asked participants to report on various components of their SNS use: the frequency with which they visited specific SNSs (ie, weekly frequency), time per day, habit, engagement, enjoyment, and addiction. We also assessed the average emotional impact of discrete SNS activities (eg, looking at memes) as well as how much one’s SNS content to which they were exposed was related to COVID-19 after the outbreak. How we measured these SNS components is described in detail below.

**Weekly Frequency**

We assessed weekly frequency of SNS use by presenting participants with a list of seven SNSs: Facebook, Instagram, Twitter, Snapchat, Reddit, Tumblr, and LinkedIn. These sites were selected based on two selection criteria: sites on which the people in one’s network are people whom one is likely to know “in real life” and/or there is a significant focus on both consuming and commenting on content. Therefore, sites on which followers are unlikely to know one another in real life and on which there is not a significant focus on commenting on content (eg, TikTok) were not included. In addition, sites that are strictly text or communication based (eg, Facebook Messenger) were also not included. Participants endorsed how frequently they used each of the seven sites in a typical week of the given time frame (ie, weekly frequency). These items were scored on an experimenter-generated 8-point Likert scale from 1 (never) to 8 (7+ times per day). Values were summed across the seven SNSs, such that total weekly frequency scores could range from 7 to 56.

iPhones have a Screen Time function in phone settings that provides a breakdown of cell phone use activity, including average daily time spent on one’s phone and weekly total screen time; a comparable feature is not available on Android or other mobile cellular devices. Those with iPhones (156/176, 88.6%) reported these two values. Our weekly frequency variable was positively correlated with participants’ “weekly total screen time” on their iPhones ($r=0.26, P=.002$), suggesting that participants were able to accurately estimate how much they visit their phones (and SNSs) each week.

**Time per Day**

To determine how much time participants spent on each of the seven SNSs (ie, time per day), they were directly asked to report “how much time in a typical day” in the given time frame they had used each of the sites. For each participant, the total minutes endorsed for each of the seven sites were summed to compute a total time per day score. Our total time per day variable was significantly positively correlated with iPhone reports of “average daily time” ($r=0.41, P<.001$), suggesting that participants were able to accurately estimate how much time they spend on their phones each day.

**Habit, Engagement, and Enjoyment**

Habitual SNS use, engagement with SNSs, and enjoyment on SNSs were each assessed using the corresponding subscales developed by Turel and Serenko. For each subscale, we modified wording to refer to use across all “social networking platforms” rather than to address one specific site (eg, “Using social networking platforms has become automatic to me”). Participants were asked, “During [time frame], to what extent did you agree with the following statements?” Participants endorsed each item using a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). These three subscales have been validated on college student samples [12] and are described below.

Habitual SNS use (ie, habit) was assessed by three items: “Using social networking platforms has become automatic to me,” “Using social networking platforms is natural to me,” and
“When I want to interact with friends and relatives, using social networking platforms is an obvious choice for me.” The three values were averaged to compute a habit score. Internal consistency scores for habit were good (pre–COVID-19: α=.76; during COVID-19: α=.85).

Engagement with SNSs (ie, engagement) was assessed by three reverse-coded items: “It would not matter to me if I never used social networking platforms again,” “The less I have to do with social networking platforms, the better,” and “Social networking platforms are unimportant in my life.” The three values were averaged to compute an engagement score. Internal consistency scores for engagement were good (pre–COVID-19: α=.80; during COVID-19: α=.85).

Enjoyment on SNSs (ie, enjoyment) was assessed by five items: “Using social networking platforms is enjoyable,” “Using social networking platforms is fun,” “Using social networking platforms is exciting,” and “Using social networking platforms is interesting.” The five values were averaged to compute an enjoyment score. Internal consistency scores for enjoyment were good (pre–COVID-19: α=.86; during COVID-19: α=.89).

Addiction
SNS addiction (ie, addiction) was assessed using an adapted version of the Bergen Facebook Addiction Scale [33] that was originally developed by Shensa et al [34]. Using a 5-point Likert scale from 0 (very rarely) to 4 (very often), participants indicated the frequency with which they experienced the following six symptoms: “Spent a lot of time thinking about social networking platforms or planned use of social networking platforms,” “Felt an urge to use social networking platforms more and more,” “Used social networking platforms in order to forget about personal problems,” “Tried to cut down on your use of social networking platforms without success,” “Become restless or troubled if you have been prohibited from using social networking platforms,” and “Used social networking platforms so much that it had a negative impact on your job/studies.” Items were summed and could range from 0 to 24. This scale has been validated with a nationally representative young adult (aged 19-32 years) sample [34]. The items in this scale had good internal consistency for both administrations (pre–COVID-19: α=.89; during COVID-19: α=.89).

Average Emotional Impact From SNS Activities
To assess emotional outcomes specifically resulting from SNS use, participants were additionally presented with a 45-item list of discrete SNS activities (eg, “Read or watched news with content that I found negative or upsetting” and “Commented positively or supportively on other’s post(s)”). These items were developed through informal undergraduate focus groups and experimenter-generated items. When applicable, parallel activities were developed for items such that each activity included a positive, negative, and neutral valence (eg, “Shared a post(s) about positive events or emotions,” “Shared a post(s) about negative events or emotions,” and “Shared a post(s) about neutral (neither positive nor negative) events or emotions”). The list was presented in a random order for each participant. For each activity, participants were first asked to indicate whether they had engaged in each activity during the given time frame. For all activities endorsed, participants were then asked to indicate “what impact each of these activities had on your emotions, on average” during the given time frame on a 7-point Likert scale from 1 (made me feel really bad) to 7 (made me feel really good). These scores were summed and divided by the total number of activities the participant endorsed to calculate an average emotional impact score for each person at each time frame. Of note, individuals’ average emotional impact was significantly positively associated with enjoyment on SNSs both pre–COVID-19 (r=0.35, P<.001) and during COVID-19 (r=0.29, P<.001), suggesting that our emotional impact variable was able to adequately capture the emotional influence of SNS activities. It is important to note that this variable can also be thought of as a form of emotional well-being, although it is largely utilized as a predictor variable in this study.

COVID-19 SNS Content
To assess COVID-19–related content, participants were administered one experimenter-generated question about the extent to which their SNS content was related to COVID-19. They were asked, “Since the end of spring break (March 23rd), what percentage of your SNS content would you estimate is COVID-19 related?” Participants reported what percentage they felt their SNS content was pandemic related in a text box.

Pandemic-Related Social and Emotional Well-being Measures
We administered two additional experimenter-generated measures to assess social and emotional well-being specific to the period of time during the COVID-19 pandemic: change in perceived social support and pandemic-related stress.

Change in Perceived Social Support
We operationalized pandemic-related social well-being as “change in perceived social support” since the onset of COVID-19. It was assessed with two items: “Prior to the COVID-19 outbreak, how supported do you feel by your social network (eg, friends and family)?” and “Currently, how supported do you feel by your social network?” Participants used a 7-point Likert scale from 1 (none) to 7 (very much supported) to report the extent to which they felt socially supported at each time frame. Each participant’s score for perceived social support prior to the pandemic was subtracted from their score for current perceived social support to create the variable “change in perceived social support,” such that higher values indicate increased perceived support from pre–COVID-19 to during COVID-19. Notably, our “change in perceived social support” variable was significantly negatively associated with loneliness (r=−0.22, P=.003), lending support to the notion that this change score adequately captured the extent to which participants felt their social support had changed during COVID-19.

Pandemic-Related Stress
We operationalized pandemic-related emotional well-being as pandemic-related stress, which was assessed with two questions: “In general, what is the level of distress you have experienced with COVID-19 related to social disruptions?” and “What is your overall level of stress related to the COVID-19 outbreak?” These questions were scored on a 7-point Likert scale from 1
(no distress or no impact) to 7 (extreme distress or extreme impact). These two questions were significantly positively associated in our sample (r=0.71, P<.001), lending support to the validity of this construct of pandemic-related stress. The internal consistency for items on this scale was good (α=.83).

**Psychological Distress Measures**

**Loneliness**

To assess loneliness, we administered the UCLA (University of California, Los Angeles) Loneliness Questionnaire [35], which is a 20-item self-report scale. Participants were asked to indicate how often each of the statements is descriptive of them (eg, “I feel isolated from others”). For the purposes of this study, one item—“I find myself waiting for people to call me” —was modified to reflect more current communication practices: “I find myself waiting for people to call, text, message or otherwise contact me.” Responses were recorded on a scale from 0 (I never feel this way) to 3 (I often feel this way) and were summed. This scale was validated with a college student sample [35], and the internal consistency for items in the scale was excellent in this study’s sample (α=.95).

**General Distress**

To measure general emotional distress not necessarily attributable to the COVID-19 pandemic, we administered the Anhedonic Depression scale from the Mood and Anxiety Symptom Questionnaire (MASQ-AD). The MASQ-AD is a 22-item self-report scale that measures depressive symptomology [36]. Participants are presented with 22 items representing feelings, sensations, problems, and experiences (eg, “Felt like nothing was very enjoyable”) and are asked to report the extent to which they have experienced each item in the past week, using a 5-point Likert scale from 1 (not at all) to 5 (extremely). Items are summed to compute a total depression score. Due to the study taking place online and resulting ethical considerations, we omitted the suicidal ideation item, bringing the total number of items to 21 and the highest possible score to 105 instead of 110. Of note, this scale has been validated with three student samples and an adult sample [35], and the internal consistency for items in the scale was excellent in this study’s sample (α=.95).

**Social Anxiety**

To measure a form of social distress not necessarily attributable to the COVID-19 pandemic, we administered the Social Interaction Anxiety Scale (SIAS) [37]. The SIAS is a 21-item self-report scale that measures anxiety specific to social interactions (eg, “I have difficulty talking with other people”). Participants indicated “the degree to which you feel the statement is characteristic of you” on a 5-point Likert scale from 0 (not at all characteristic or true of me) to 4 (extremely characteristic or true of me). Scores were summed, with possible scores ranging from 0 to 84. This scale was validated with college student, community, and clinical samples [37]. The internal consistency for items in this scale in this sample was good (α=.86).

**Results**

**Sample**

A total of 183 participants were recruited from undergraduate psychology courses at a private university in the Midwestern United States to participate in a study on emotions and social media. The final sample of 176 excluded 16 individuals who did not complete any of the measures in this study. Participant ages ranged from 18 to 23 years (mean 20.00, SD 1.26). Out of 176 participants, 54.0% (n=95) identified as women and 4.5% (n=8) identified as Hispanic or Latinx. With regard to race, our participants identified as follows: 44.9% (n=79) White, 26.7% (n=47) Asian, 19.9% (n=35) Black, and 8.5% (n=15) multi-racial.

**Analytic Overview**

First, we provided descriptive statistics for each component of SNS use (ie, weekly frequency, time per day, habit, engagement, enjoyment, addiction, average emotional impact, and COVID-19–related SNS content) both before and during COVID-19, as well as pandemic-related social well-being (ie, changes in perceived social support), pandemic-related emotional well-being (ie, pandemic-related stress), and the three forms of psychological distress (ie, loneliness, depression, and social anxiety). We also presented Pearson zero-order correlations between the components of SNS use at both time frames (ie, pre– and during COVID-19). Then to assess effects of gender and race, we conducted a factorial multivariate analysis of variance, such that the components of SNS use were predicted by race and gender across the two time frames.

Aim 1 was to examine how SNS use has changed from pre–COVID-19 to during COVID-19, by comparing the means of the seven components of SNS use from pre–COVID-19 to during COVID-19 via a series of paired-sample t tests. We did not examine COVID-19–related SNS content because this construct was only assessed once during COVID-19.

Aim 2 was to examine how changes in components of SNS use during the pandemic were related to pandemic-related social well-being (ie, change in perceived social support) and pandemic-related emotional well-being (ie, pandemic-related stress). We created a residualized variable for each component of SNS use for which we assessed change, with each of the resulting variables representing the component of SNS use during COVID-19 that cannot be explained or predicted by the same component of SNS use pre–COVID-19; we will call this “change in SNS use components.” We conducted Pearson correlations between the change in SNS use components as well as COVID-19 SNS content and pandemic-related social and emotional well-being. Then, we conducted two linear regressions where we simultaneously entered the change in SNS use components and general distress (ie, depression) to predict pandemic-related social and emotional well-being. This allowed us to examine which changes in SNS use were uniquely related to the two outcomes while controlling for general distress.

Finally, Aim 3 was to examine how loneliness was associated with components of SNS use during the pandemic specifically and with change in SNS use from pre–COVID-19 to during the
COVID-19 pandemic. First, we conducted zero-order Pearson correlations between the eight SNS components during COVID-19 and loneliness. Then, to examine unique effects of the SNS use components during COVID-19 on loneliness, we entered the eight SNS components simultaneously to predict loneliness. We also included social anxiety as a covariate so that effects were specific to loneliness and were not better explained by social anxiety. Next, to assess how changes in SNS use during COVID-19 are related to loneliness, we conducted Pearson correlations between loneliness and the seven “change in SNS use components” as well as COVID-19 SNS content. Finally, to assess unique effects of changes in SNS use on loneliness, we conducted a linear regression in which we entered the changes in SNS use components and COVID-19 SNS content and social anxiety simultaneously to predict loneliness.

Descriptive Statistics and Correlations

Descriptive statistics for the eight components of SNS use for both time frames are presented in Table 1. Overall, these values are similar to existing work utilizing student samples [11,23]. With regard to descriptive statistics for the psychological distress measures, loneliness was lower than would be expected in a university sample (mean 23.09, SD 13.63) [24,27]. Levels of depressive symptoms were similar to other student samples (mean 56.76, SD 11.02) [36], and the sample can be characterized as having low levels of depression [25,28], although there was a wide range of values, including some above established clinical cutoff values [38]. Social anxiety was higher than typical in a student sample, but lower than would be expected in clinical samples (mean 33.80, SD 12.52) [26,29,30], indicating somewhat moderate levels of social anxiety in our sample, on average.

Zero-order Pearson correlations between the SNS use components pre– and during COVID-19 are presented in Table 2. Pre–COVID-19, correlations between SNS components ranged from –0.07 to 0.51 (mean 0.22, SD 0.17). During COVID-19, correlations between SNS components ranged from –0.13 to 0.56 (mean 0.25, SD 0.19). Test-retest correlations between time frames for each of the seven SNS components were generally large, ranging from 0.66 to 0.88 (mean 0.77, SD 0.07). There were not significant effects of gender ($F_{1,163}=1.48$, $P=.12$) or race ($F_{3,163}=0.94$, $P=.59$) on any of the components of SNS use.
Table 1. Descriptive statistics for social networking site (SNS) use pre–COVID-19 and during COVID-19.

<table>
<thead>
<tr>
<th>Components of SNS use</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre–COVID-19</strong></td>
<td></td>
</tr>
<tr>
<td>Weekly frequency(^a)</td>
<td>24.12 (6.60)</td>
</tr>
<tr>
<td>Time per day (minutes)(^b)</td>
<td>115.83 (113.53)</td>
</tr>
<tr>
<td>Habit(^c)</td>
<td>3.95 (0.82)</td>
</tr>
<tr>
<td>Enjoyment(^c)</td>
<td>3.61 (0.65)</td>
</tr>
<tr>
<td>Engagement(^c)</td>
<td>3.44 (0.93)</td>
</tr>
<tr>
<td>Addiction(^c)</td>
<td>8.98 (5.40)</td>
</tr>
<tr>
<td>Average emotional impact(^d)</td>
<td>4.29 (0.38)</td>
</tr>
<tr>
<td><strong>During COVID-19</strong></td>
<td></td>
</tr>
<tr>
<td>Weekly frequency(^a)</td>
<td>24.57 (7.41)</td>
</tr>
<tr>
<td>Time per day (minutes)(^b)</td>
<td>196.38 (162.33)</td>
</tr>
<tr>
<td>Habit(^c)</td>
<td>4.11 (0.87)</td>
</tr>
<tr>
<td>Enjoyment(^c)</td>
<td>3.53 (0.77)</td>
</tr>
<tr>
<td>Engagement(^c)</td>
<td>3.52 (1.04)</td>
</tr>
<tr>
<td>Addiction(^c)</td>
<td>10.55 (6.02)</td>
</tr>
<tr>
<td>Average emotional impact(^d)</td>
<td>4.18 (0.50)</td>
</tr>
<tr>
<td>COVID-19 SNS content (%)(^e)</td>
<td>42.57 (22.89)</td>
</tr>
</tbody>
</table>

\(^a\) Items were scored on an experimenter-generated 8-point Likert scale from 1 (never) to 8 (7+ times per day). Summed scores could range from 7 to 56.

\(^b\) Participants were directly asked to report “how much time in a typical day” in the given time frame they had used each of the seven SNSs.

\(^c\) Items were scored on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). Averaged scores could range from 1 to 5.

\(^d\) For each activity, participants indicated whether they had engaged in it during the given time frame. For all activities endorsed, participants indicated what impact each had on their emotions, on average, during the given time frame on a 7-point Likert scale from 1 (made me feel really bad) to 7 (made me feel really good). Scores were summed and divided by the total number of activities the participant endorsed.

\(^e\) Participants reported what percentage they felt their SNS content was pandemic related in a text box.
Table 2. Correlation analysis (Pearson zero-order *r* and two-tailed *P* value) among social networking site (SNS) components pre–COVID-19 and during COVID-19.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Weekly frequency</th>
<th>Time per day</th>
<th>Habit</th>
<th>Enjoyment</th>
<th>Engagement</th>
<th>Addiction</th>
<th>Average emotional impact</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Weekly frequency pre–COVID-19</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>r</em></td>
<td>1</td>
<td>0.42&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.43&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.24&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.29&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.36&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.01</td>
</tr>
<tr>
<td><em>P</em> value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Weekly frequency during COVID-19</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>r</em></td>
<td>1</td>
<td>0.41&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.34&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.25&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.30&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.25&lt;sup&gt;a&lt;/sup&gt;</td>
<td>–0.07</td>
</tr>
<tr>
<td><em>P</em> value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Weekly frequency between pre– and during COVID-19</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>r</em></td>
<td>0.88&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P</em> value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time per day pre–COVID-19</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>r</em></td>
<td></td>
<td>0.42&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>0.26&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.19&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.09&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.30&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>P</em> value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time per day during COVID-19</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>r</em></td>
<td></td>
<td>0.41&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>0.25&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.17&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.22&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.19&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>P</em> value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time per day between pre– and during COVID-19</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>r</em></td>
<td></td>
<td></td>
<td>0.84&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P</em> value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Habit pre–COVID-19</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>r</em></td>
<td></td>
<td></td>
<td>0.43&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.26&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>0.42&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.56&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>P</em> value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Habit during COVID-19</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>r</em></td>
<td></td>
<td></td>
<td>0.34&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.25&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>0.46&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.51&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>P</em> value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Habit between pre– and during COVID-19</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>r</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P</em> value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Enjoyment pre–COVID-19</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>r</em></td>
<td></td>
<td></td>
<td>0.24&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.19&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.42&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>0.51&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>P</em> value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Enjoyment during COVID-19</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>r</em></td>
<td></td>
<td></td>
<td>0.25&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.17&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.46&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>0.52&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>P</em> value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Enjoyment between pre– and during COVID-19</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>r</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P</em> value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Engagement pre–COVID-19</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>r</em></td>
<td></td>
<td></td>
<td>0.29&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.09&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.56&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.51&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
</tr>
</tbody>
</table>
Aim 1. How Components of SNS Use Changed From Before to During COVID-19

Consistent with our hypothesis, there was an increase in daily time spent on SNSs ($t_{169}=5.53$, $d=0.42$, $P<.001$), habitual use of SNSs ($t_{169}=3.60$, $d=0.27$, $P<.001$), and SNS addiction ($t_{173}=4.96$, $d=0.38$, $P<.001$) during COVID-19 compared to pre–COVID-19. In addition, the average impact of endorsed SNS activities on emotions became more negative ($t_{172}=–3.76$, $d=–0.29$, $P<.001$). Inconsistent with our hypotheses, enjoyment on SNSs decreased ($t_{172}=–2.10$, $d=–0.16$, $P=.04$). Weekly frequency of SNS use ($t_{174}=1.74$, $d=0.14$, $P=.08$) and engagement with SNSs ($t_{174}=1.53$, $d=0.12$, $P=.13$) did not significantly change during COVID-19.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Weekly frequency</th>
<th>Time per day</th>
<th>Habit</th>
<th>Enjoyment</th>
<th>Engagement</th>
<th>Addiction</th>
<th>Average emotional impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
<td>.35</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>—</td>
<td>.09</td>
<td>.006</td>
</tr>
</tbody>
</table>

Engagement during COVID-19

$r$ 0.30$^a$ 0.22$^a$ 0.51$^a$ 0.52$^a$ 1 0.23$^a$ 0.13

$p$ value <.001 <.001 <.001 <.001 <.001 — <.001

Engagement between pre– and during COVID-19

$r$ — — — — .79$^a$ — —

$p$ value — — — — <.001 — —

Addiction pre–COVID-19

$r$ 0.36$^a$ 0.30$^a$ 0.33$^a$ 0.19$^a$ 0.13 1 −0.13

$p$ value <.001 <.001 <.001 .01 .09 — .09

Addiction during COVID-19

$r$ 0.25$^a$ 0.19$^a$ 0.32$^a$ 0.09 0.23$^a$ 1 —

$p$ value <.001 .03 <.001 .26 .003 — —

Addiction between pre– and during COVID-19

$r$ — — — — — 0.72$^a$ —

$p$ value — — — — — <.001 —

Average emotional impact pre–COVID-19

$r$ 0.01 0.03 0.22$^a$ 0.29$^a$ 0.21$^a$ −0.13 1

$p$ value .90 .61 .004 <.001 .006 .09 —

Average emotional impact during COVID-19

$r$ −0.07 −0.04 0.06 0.35$^a$ 0.13 −0.26$^a$ 1

$p$ value .36 .88 .56 <.001 .06 <.001 —

Average emotional impact between pre– and during COVID-19

$r$ — — — — — — 0.66$^a$

$p$ value — — — — — — <.001

$^a$The correlation is significant at a significance level of .05 (two-tailed).

$^b$Not applicable.

Aim 2. How Changes in Components of SNS Use Are Related to Pandemic-Related Social and Emotional Well-being

Change in Perceived Social Support

First, consistent with our hypothesis, we found that change in weekly frequency, change in time per day, and change in addiction were each positively associated with increased social support. Contrary to our hypothesis, change in engagement was also positively associated with increased social support (Table 3). That is, compared to pre–COVID-19, those who visited SNSs more frequently during COVID-19, those who spent more time on SNSs during COVID-19, those who experienced more SNS addiction during COVID-19, and those who were more engaged on SNSs during COVID-19 endorsed greater increases in perceived social support during COVID-19. Inconsistent with our hypothesis, changes in SNS habit, average emotional impact, and COVID-19 SNS content were not significantly associated...
with perceived change in social support during the pandemic. When examining all seven “change in SNS use components,” COVID-19 SNS content, and depression in the same model predicting change in perceived social support, no associations were significant. These findings suggest that no changes in SNS use from pre- to during COVID-19 uniquely predict change in perceived social support.

Table 3. Pearson zero-order correlations between changes in social networking site (SNS) use components and pandemic-related social and emotional well-being measures.

<table>
<thead>
<tr>
<th>Variable</th>
<th>$r$</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Change in perceived social support</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change in weekly frequency</td>
<td>0.24$^b$</td>
<td>.007</td>
</tr>
<tr>
<td>Change in time per day (minutes)</td>
<td>0.20$^b$</td>
<td>.02</td>
</tr>
<tr>
<td>Change in habit</td>
<td>0.15</td>
<td>.08</td>
</tr>
<tr>
<td>Change in enjoyment</td>
<td>0.10</td>
<td>.30</td>
</tr>
<tr>
<td>Change in engagement</td>
<td>0.20$^b$</td>
<td>.02</td>
</tr>
<tr>
<td>Change in addiction</td>
<td>0.18$^b$</td>
<td>.02</td>
</tr>
<tr>
<td>Change in average emotional impact</td>
<td>0.04</td>
<td>.46</td>
</tr>
<tr>
<td>COVID-19 SNS content</td>
<td>−0.01</td>
<td>.68</td>
</tr>
<tr>
<td><strong>Pandemic-related stress</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change in weekly frequency</td>
<td>0.06</td>
<td>.33</td>
</tr>
<tr>
<td>Change in time per day (minutes)</td>
<td>0.07</td>
<td>.38</td>
</tr>
<tr>
<td>Change in habit</td>
<td>0.05</td>
<td>.48</td>
</tr>
<tr>
<td>Change in enjoyment</td>
<td>−0.02</td>
<td>.79</td>
</tr>
<tr>
<td>Change in engagement</td>
<td>0.12</td>
<td>.13</td>
</tr>
<tr>
<td>Change in addiction</td>
<td>0.23$^b$</td>
<td>.002</td>
</tr>
<tr>
<td>Change in average emotional impact</td>
<td>0.07</td>
<td>.39</td>
</tr>
<tr>
<td>COVID-19 SNS content</td>
<td>0.20$^b$</td>
<td>.006</td>
</tr>
</tbody>
</table>

$^a$These analyses utilizing change components used residualized SNS use variables.

$^b$The correlation is significant at a significance level of .05 (two-tailed).

**Pandemic-Related Stress**

Consistent with our hypothesis, we found that pandemic-related stress was significantly positively associated with change in addiction and COVID-19 SNS content. Inconsistent with our hypothesis, pandemic-related stress was not associated with changes in frequency, time, habit, engagement, or average emotional impact of SNS use (Table 3). When predicting pandemic-related stress from the seven “change in SNS use components,” COVID-19 SNS content, and general distress (covariate), these results showed the same pattern of findings (addiction: $\beta=0.07$, $t_{159}=2.90$, $P=.004$; COVID-19 SNS content: $\beta=0.01$, $t_{159}=2.44$, $P=.02$). These findings suggest that both increased addictive SNS use and the percentage of one’s SNS content related to the pandemic are associated with pandemic-related stress, even after taking into account the other SNS use components and general distress.

**Loneliness**

We assessed how loneliness is associated with SNS use during the pandemic. Consistent with our hypothesis, we found that loneliness was positively associated with addiction during COVID-19 and negatively associated with engagement and the average emotional impact of SNSs during COVID-19. Inconsistent with our hypotheses, loneliness was not significantly associated with weekly frequency, time per day, habit, enjoyment, or COVID-19 SNS content (Table 4). However, when examining all eight SNS components during COVID-19 and social anxiety (as a covariate) simultaneously, only addiction was significant ($\beta=0.50$, $t_{159}=2.78$, $P=.006$). These results suggest that it is only addictive SNSs that are uniquely associated with loneliness during COVID-19.

Second, we assessed how loneliness is associated with changes in SNS use during the pandemic. Consistent with our hypothesis, we found that loneliness was negatively associated with change in engagement. That is, those who were lonelier endorsed less SNS engagement during COVID-19 compared to their endorsed engagement pre–COVID-19. Contrary to our hypothesis, loneliness was also negatively associated with change in habit. Also inconsistent with hypotheses, loneliness was not significantly associated with changes in weekly frequency, time per day, engagement, addiction, average emotional impact, or COVID-19 SNS content (Table 4). When we considered the
seven “change in SNS use components” simultaneously while controlling for social anxiety, change in habit was significant ($\beta=-4.86, t_{158}=-2.67, P=.001$), engagement was not significant ($\beta=-2.11, t_{158}=-1.29, P=.20$), and addiction was significant ($\beta=0.49, t_{158}=1.98, P=.049$). In this way, addiction was revealed in the linear regression model as a previously suppressed variable that required a more powerful test with decreased standard error to be revealed. These results suggest that loneliness was uniquely associated with using SNSs less habitually and being more addicted to SNSs during the pandemic.

Table 4. Pearson zero-order correlations between loneliness and social networking site (SNS) use components during COVID-19 and changes in SNS use components from pre– to during COVID-19.

<table>
<thead>
<tr>
<th>Variable</th>
<th>$r$</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SNS use during COVID-19</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change in weekly frequency</td>
<td>0.07</td>
<td>.34</td>
</tr>
<tr>
<td>Change in time per day (minutes)</td>
<td>0.10</td>
<td>.19</td>
</tr>
<tr>
<td>Change in habit</td>
<td>-0.07</td>
<td>.17</td>
</tr>
<tr>
<td>Change in enjoyment</td>
<td>-0.10</td>
<td>.18</td>
</tr>
<tr>
<td>Change in engagement</td>
<td>-0.16b</td>
<td>.04</td>
</tr>
<tr>
<td>Change in addiction</td>
<td>0.26b</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Change in average emotional impact</td>
<td>-0.19b</td>
<td>.01</td>
</tr>
<tr>
<td>COVID-19 SNS content</td>
<td>0.07</td>
<td>.34</td>
</tr>
<tr>
<td><strong>Changes in SNS use from pre– to during COVID-19</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change in weekly frequency</td>
<td>0.07</td>
<td>.39</td>
</tr>
<tr>
<td>Change in time per day (minutes)</td>
<td>0.00</td>
<td>.99</td>
</tr>
<tr>
<td>Change in habit</td>
<td>-0.21b</td>
<td>.01</td>
</tr>
<tr>
<td>Change in enjoyment</td>
<td>-0.08</td>
<td>.27</td>
</tr>
<tr>
<td>Change in engagement</td>
<td>-0.15b</td>
<td>.04</td>
</tr>
<tr>
<td>Change in addiction</td>
<td>0.12</td>
<td>.31</td>
</tr>
<tr>
<td>Change in average emotional impact</td>
<td>0.02</td>
<td>.98</td>
</tr>
</tbody>
</table>

aThese analyses used variables assessed with the time frame of during COVID-19.
bThe correlation is significant at a significance level of .05 (two-tailed).
cThese analyses used residualized SNS use variables.

**Discussion**

This investigation examined how the COVID-19 pandemic was associated with changes in SNS use and how these changes were associated with psychological outcomes in a college student sample. This study expands upon the literature in several important ways. First, rather than assessing only frequency of SNS use, we examined how multiple components of SNS use (ie, weekly frequency, time per day, habit, engagement, enjoyment, addiction, average emotional impact, and COVID-19-related SNS content) changed from pre– to during COVID-19, and how these changes in SNS use were related to social and emotional well-being. Second, this investigation assessed how these components of SNS use were related to loneliness during a global pandemic when rates of loneliness are believed to be elevated. Lastly, to our knowledge, this was the first investigation to examine the perceived impact of engagement in SNS activities on people’s emotions and how this was associated with the COVID-19 pandemic.

The study’s first aim was to examine how SNS use changed from pre– to during COVID-19. Mostly consistent with hypotheses based on Turel and Serenko’s [12] path model, changes in SNS use, including time spent on SNSs; habitual SNS use; and SNS addiction increased, while enjoyment on SNSs decreased and the average emotional impact of SNS activities became more negative. These findings suggest that individuals’ increased thoughts and behaviors toward SNSs during COVID-19 (ie, spending more time on SNSs, using them more habitually, and experiencing greater addiction) could be maladaptive and are associated with more negative emotional experiences.

The study’s second aim was to investigate how changes in components of SNS use during the pandemic were related to pandemic-related social and emotional well-being. Contrary to our hypotheses that increased SNS use would be negatively associated with pandemic-related social and emotional well-being, greater increases in perceived social support during COVID-19 were associated with (1) more frequent SNS use, (2) more time spent on SNSs, (3) greater SNS addiction, and
(4) greater engagement with SNSs during COVID-19. However, when all SNS components were taken into consideration, none were significantly associated with perceived social support. This suggests that no one way of using SNSs (ie, using them more frequently, more addictively, etc) uniquely accounted for increased perceptions of social support during COVID-19. Consequently, results should be interpreted with caution. Nonetheless, these findings provide some evidence that SNS use during the pandemic could be socially adaptive and might create a space for individuals to feel more socially connected. SNS addiction during COVID-19 and the extent to which one’s SNS content was related to the pandemic were associated with greater pandemic-related stress, controlling for general distress, consistent with hypotheses.

Overall, the associations between components of SNS use and pandemic-related well-being were mixed. Greater perceived social support during COVID-19 was associated with using SNSs more frequently and for more time, as well as reporting greater SNS engagement and addiction. In contrast, SNS addiction during COVID-19 and exposure to COVID-19–related SNS content were each associated with decreases in pandemic-related social and emotional well-being. And compared to pre–COVID-19, individuals during COVID-19 reported enjoying SNSs less and experiencing greater negative impacts of SNS activities on their emotions. Overall, these results suggest that, despite individuals using SNSs more frequently and for more time during the pandemic, use of SNSs during COVID-19 was associated with mixed social outcomes and largely negative emotional outcomes.

The study’s third aim was to examine how loneliness was associated with SNS use during the pandemic. Consistent with hypotheses, higher levels of loneliness were significantly associated with SNS activities during COVID-19 having a negative emotional impact. Loneliness was positively associated with SNS addiction during COVID-19 and negatively associated with engagement in SNSs during COVID-19. Importantly, addictive SNS use during COVID-19 was significantly related to loneliness even after accounting for the other SNS use components and social anxiety. Inconsistent with hypotheses, loneliness was associated with reductions in habitual SNS use and engagement on SNSs from pre– to during COVID-19. However, when simultaneously considering how all SNS components and social anxiety were associated with loneliness, only reductions in habit remained significant, and addictive SNS use became significant.

Although these findings illustrate associations between loneliness and various components of SNS use, further research is needed to determine directionality between these constructs. On the one hand, it is possible that people were lonely during the pandemic because they were not using SNSs as habitually as they once had. On the other hand, it could be that those who were lonely during COVID-19 were aware of the negative impact of SNS use on their emotions and mental health and, therefore, chose to engage with SNSs less habitually during the pandemic. In a sense, a decrease in SNS habit during COVID-19 could serve as a protective mechanism for those high in loneliness. Although findings showed that decreases in habit were associated with loneliness, increases in SNS addiction during the pandemic were also associated with loneliness. Perhaps individuals who were lonely during COVID-19 stopped using SNSs habitually and used them more addictively instead, an outcome that may occur if individuals wish to use SNSs less but still find themselves turning to them.

Our additional findings that loneliness was associated with SNS activities having a more negative, or less positive, emotional impact may shed important insight into the role of SNSs on loneliness. Research suggests that loneliness causes increased SNS use, and not that SNS use causes increased loneliness [21]. Despite lonely people using SNSs more than others, increased loneliness is also correlated with experiencing greater negative influences of SNSs on emotions. Again, further research is needed to understand the temporal relations between these constructs. It could be that when individuals who are lonely turn to SNSs to receive social stimuli, (1) this exposure leads to social comparison experiences that result in negative emotions (ie, seeing pictures of peers spending time together) or (2) they may not reap the same emotional benefits from them as do those who are less lonely. Future research should examine these possibilities to begin to elucidate how emotions for those who are lonely are implicated in SNS use.

Interestingly, across analyses examining our three aims, addictive SNS use was the SNS activity that was most consistently significant in our models, highlighting its potential importance in predicting well-being. Namely, addiction significantly increased from pre– to during COVID-19. Addiction was also associated with increases in perceived change in social support during COVID-19, greater pandemic-related stress, and greater loneliness. These findings suggest that SNS users should be aware of their addictive SNS tendencies and be cognizant of how this addictive use may be associated with their well-being (eg, noting that addictive SNS use makes them feel more socially connected, but also more stressed). Future research should continue to explore the role of SNS addiction in individuals’ everyday lives and emotional experiences, especially during times of global health crises when SNS use seems to increase.

This investigation highlights potential clinical implications. During COVID-19, therapy clients, like this study’s sample, may report that their use of SNSs during COVID-19 has increased. In these cases, it may be helpful for mental health providers to note that increased SNS use during COVID-19 has been linked to mixed outcomes, at least in a college student sample. Mental health providers could help their clients examine when SNS use may be adaptive versus maladaptive and when clients, for example, should pursue other social outlets (eg, having in-person conversations). This is consistent with cognitive behavioral therapies, which place emphasis on helping clients engage in behaviors that have an emotion-boosting effect and limiting behaviors that negatively influence emotions [39]. Examining SNS use in therapy seems particularly critical for clients who report elevated loneliness, since they may be using SNSs more than others.

Furthermore, these findings suggest utility in assessing and monitoring for SNS addiction specifically. Although SNS addiction, as assessed in this study, was associated with
increased perceived social support, it was also associated with greater stress and loneliness. An important avenue for future research is to examine at which levels SNS addiction causes clinically significant distress or impairment, which are requirements for receiving formal diagnoses of addictive disorders [40]. It will also be important to establish which treatments are best suited for treating SNS addiction since there are not currently any empirically supported treatments for it. Of note, experts have posited that treatment should center on controlling SNS use, rather than abstaining from it, since SNSs have become an integral and unavoidable part of life [39,40], which may be especially true for college-aged individuals. In addition, it might be useful to provide psychoeducation to clients about SNS addiction.

This investigation has several limitations. Most notably, since data were collected about one month into the pandemic, it is not known whether these trends of SNS use have continued throughout the pandemic. However, given the initial surge in use in this sample, we would expect trends in SNS use to persist on the premise that—as can be seen in this investigation as well as others—SNS use is habitual [12]. Both SNS addiction and using SNSs as a means to connect with others are associated with habitual SNS use [12,41]. Therefore, given the increase in SNS addiction seen in this sample and our theorizing that SNSs were used as tools for social connection, it is not surprising that our data showed an increase in habit near the start of the pandemic. Once habits are formed, they are very difficult to break [42], leading us to think that these trends in SNS use witnessed at the beginning of the pandemic would remain today.

It is also important for SNS research to utilize designs other than retrospective reports, which can be biased and more difficult for participants to accurately complete [43]. For example, prospective longitudinal research could be utilized to examine how SNS use and its associations with well-being evolve over the course of disease outbreaks. Additionally, because this investigation focused on college students, the findings may not generalize to peers who do not attend college or to older samples. Consequently, this study did not shed light on how SNS use during disease outbreaks is related to loneliness and various indices of well-being across the adult lifespan. We expect that use of SNSs during the pandemic would be related to positive outcomes, particularly in older adults. Older adults have been shown to have generally positive feelings toward SNSs, and SNS use in this population is associated with greater well-being and less loneliness [43,44].

In a world where an increasing amount of time and social interactions are occurring in an online sphere, it is imperative to investigate and understand the role of SNSs on social and emotional well-being during times of crisis. Findings from this investigation highlight both benefits and disadvantages to SNS use, underscoring the nuanced and multifaceted nature of the correlates of these sites with well-being. Although the COVID-19 pandemic may be one of the first globally salient incidents that has erupted since the widespread adoption of SNS use, it is unlikely to be the last. It is hoped that findings from this investigation will advise SNS users on how to best cope with the COVID-19 pandemic and any future pandemics as well. This study and those like it are only beginning to help us truly understand how SNS use is associated with our everyday social and emotional well-being during stressful and trying times.

Acknowledgments

We thank Daphne Liu and Dr Mike Strube for their feedback on the manuscript.

Conflicts of Interest

None declared.

References


Abbreviations

MASQ-AD: Anhedonic Depression scale from the Mood and Anxiety Symptom Questionnaire
SIAS: Social Interaction Anxiety Scale
SNS: social networking site
UCLA: University of California, Los Angeles
Domestic Violence and Mental Health During the COVID-19 Pandemic in Bangladesh

Tanjir Rashid Soron¹, MBBS, MPH, MSc, MD; Md Ashiqur Rahman Ashiq¹, MBBS, MPH; Marzia Al-Hakeem¹, MS, MPhil; Zaid Farzan Chowdhury¹, BSc; Helal Uddin Ahmed², MBBS, MD; Chaman Afrooz Chowdhury³, MBBS, MPhil

¹Telepsychiatry Research and Innovation Network Ltd, Dhaka, Bangladesh
²Department of Child, Adolescent, and Family Psychiatry, National Institute of Mental Health, Dhaka, Bangladesh
³Sir Salimullah Medical College Hospital, Dhaka, Bangladesh

Corresponding Author:
Tanjir Rashid Soron, MBBS, MPH, MSc, MD
Telepsychiatry Research and Innovation Network Ltd
3D, 2nd Floor
Rupayan Trade Center, Bangla Motor
Dhaka, 1205
Bangladesh
Phone: 880 1718827138
Email: tanjirsoron@gmail.com

Abstract

Background: The COVID-19 lockdown, the advent of working from home, and other unprecedented events have resulted in multilayer and multidimensional impacts on our personal, social, and occupational lives. Mental health conditions are deteriorating, financial crises are increasing in prevalence, and the need to stay at home has resulted in the increased prevalence of domestic violence. In Bangladesh, where domestic violence is already prevalent, the lockdown period and stay-at-home orders could result in more opportunities and increased scope for perpetrators of domestic violence.

Objective: In this study, we aimed to determine the prevalence and pattern of domestic violence during the initial COVID-19 lockdown period in Bangladesh and the perceptions of domestic violence survivors with regard to mental health care.

Methods: We conducted this cross-sectional web-based study among the Bangladeshi population and used a semistructured self-reported questionnaire to understand the patterns of domestic violence and perceptions on mental health care from August to September 2020. The questionnaire was disseminated on different organizational websites and social media pages (ie, those of organizations that provide mental health and domestic violence services). Data were analyzed by using IBM SPSS (version 22.0; IBM Corporation).

Results: We found that 36.8% (50/136) of respondents had faced domestic violence at some point in their lives; psychological abuse was the most common type of violence. However, the prevalence of the economical abuse domestic violence type increased after the COVID-19 lockdown was enforced. Although 96.3% (102/136) of the participants believed that domestic violence survivors need mental health support, only 25% (34/136) of the respondents had an idea about the mental health services that are available for domestic violence survivors in Bangladesh and how and where they could avail mental health services.

Conclusions: Domestic violence is one of the most well-known stressors that have direct impacts on physical and mental health. However, the burden of domestic violence is often underreported, and its impact on mental health is neglected in Bangladesh. The burden of this problem has increased during the COVID-19 crisis, and the cry for mental health support is obvious in the country. However, it is necessary to provide information about available support services; telepsychiatry can be a good option for providing immediate mental health support in a convenient and cost-effective manner.

(JMIR Form Res 2021;5(9):e24624) doi:10.2196/24624

KEYWORDS
domestic violence; COVID-19; mental health; violence; Bangladesh; lockdown; isolation; anxiety; stress; telemental health; telepsychiatry; web-based survey
**Introduction**

A person’s home is one of the safest and most secure and beloved areas where one can dream about enjoying every moment. However, a large number of people around the world feel insecure, frightened, and panic due to experiencing physical, sexual, or psychological violence in their home environment (ie, violence from a familiar and related person). Domestic violence refers to violence that takes place in intimate relationships or when there is a relationship between the violence survivor and the perpetrator. This means that partners, ex-partners, close or distant family members, relatives, or family friends—anyone—can cause such violence. According to the Domestic Violence Prevention and Protection Act 2010, domestic violence is defined as “physical, psychological, sexual or economic abuse against a woman or child of a family by any other person of that family with whom the victim is, or has been, in family relationship.” Domestic violence, which is also sometimes referred to as intimate partner violence, is normally assumed to be perpetrated against females, but in general, domestic violence can be perpetrated against anyone in an intimate relationship. The main determinants of these kinds of behaviors may be the desire to acquire or establish a power balance and exert control over a partner or relationship. Usually, the unequal dynamics in a relation are thought to be one of the main contributing factors [1].

The World Health Organization has estimated that about 35% of women worldwide have experienced either physical and/or sexual intimate partner violence or nonpartner sexual violence in their lifetime [2]. Although domestic violence is a universal problem, sociocultural influences and the portrayal of domestic violence in the media characterize its local pattern and define the acceptance, expression, and explanation of the problem. According to the Bangladesh Bureau of Statistics, more than 70% of married women in Bangladesh have reported at least 1 physical or sexual violence incident in their conjugal life [3]. There are many unreported incidents of physical, emotional, verbal, sexual, and financial abuse. The recent COVID-19 pandemic has made gender-based violence more prominent, as the problem has deepened and has invaded new families [4]. This global health crisis has imposed multidimensional and multiphase negative impacts on health, social areas, and economic sectors. Researchers have found that gender-based violence increases during and after unprecedented humanitarian crises, including conflicts and natural disasters [4].

There are several determinants that can result in the increased prevalence of gender-based violence or domestic violence during any emergency or crisis. First, the preexisting economic dependency of females on their male counterparts intensifies the risk of gender-based violence during days of crisis [5]. Second, during a crisis situation, survivors of gender-based violence are often deprived of ample legal support, and as a result, the perpetrators remain unpunished [6]. Third, the absence of scrutiny from the outside world during a pandemic can distort the power balance at home, which can result in violence and abuse [7].

The United Nations Fund for Population Activities reported at least a 20% increase in the incidence of violence during the COVID-19 pandemic in 193 member states of the United Nations [8]. A telesurvey, which was conducted by the Manusher Jonno Foundation in 53 districts of Bangladesh, revealed that 9844 women and 2896 children experienced domestic violence until June 2020. Further, of the total number of reported cases, 4160 participants experienced such violence for the first time in their lives [5].

The COVID-19 pandemic has brought unforeseen crises such as poverty, physical distancing, and violence to the surface, and a large number of people are losing jobs or having their salary cut. Many women earn money by working as helping hands for different families in Bangladesh, and this type of daily work is the only source of income for thousands of women. However, when the daily number of people with SARS-CoV-2 infection increased, many families stopped allowing women to work in their homes due to the risk of infection. As a result, most women have started searching for food at one point or another and have become a vulnerable group that has a high risk of experiencing sexual abuse and violence, as evidenced by a couple of rape incidents that have been documented [6].

Domestic violence has resulted in long-term psychological trauma and impacts in addition to physical injuries and economic harassment. However, its threat to mental health and well-being is far more pervasive, severe, and long-term. Studies have shown that women exposed to gender-based violence have more than a twofold higher risk of developing a common mental disorder, including depression, anxiety disorder, posttraumatic stress disorder, and substance abuse, and suicidal tendencies [7]. A study conducted in Bangladesh found that reported negative health consequences range from simple injuries to grievous hurt and include psychological consequences, such as depression, anxiety, obsession, posttraumatic stress disorder, and even suicidal tendencies [9,10]. There has been an exponential rise in the prevalence of mental illnesses, including depression, anxiety, posttraumatic stress disorder, and suicidal ideation, among women who have experienced violence. Further, as the prevalence of violence against women increases, the need for mental health care also increases [11]. Moreover, there is a bidirectional relationship between mental health and gender-based violence. Women living with a severe mental illness are significantly more likely to experience violence [12]. Therefore, mental health care is an important and integral part of reducing the burden of gender-based violence. Globally, women have limited access to mental health care, and the situation is more disappointing in low- and middle-income countries like Bangladesh. In this country, 1 psychiatrist serves roughly 1 million people, and there are less than 50 clinical psychologists among the whole population [13]. The situation becomes more complicated when one considers that most of the mental health professionals and mental health care facilities in Bangladesh are located in the capital city—Dhaka.

The main purpose of this study was to obtain a view of the rates of domestic violence during the pandemic and the perceptions of domestic violence survivors with regard to mental health care.
Methods

Study Summary
This cross-sectional study was conducted via a web-based platform from August to September 2020, which was when all of Bangladesh was struggling to manage the COVID-19 crisis. Both males and females aged above 16 years were eligible to participate in this study. A structured questionnaire was designed by the authors to fulfill the objectives of this study. Before participants’ participated in this study, an information sheet and a consent form were made available (ie, on the first page of the questionnaire). They were written in the Bangla languages so that information could be easily understood by all of the participants. The participants were free to withdraw at any time without having to provide explanations and were also allowed to not answer any questions if they wished. Moreover, personal identification information was not requested from participants to maintain information confidentiality.

Design
The convenience and snowball sampling techniques were used to obtain the desired sample size. We used professional organizations’ and volunteer organizations’ social media pages to disseminate the questionnaire. In addition, we encouraged the recipients of the questionnaire to send the questionnaire to their friends and family members for completion. The questionnaire was composed of information on the sociodemographic characteristics of the respondents, the impact that the COVID-19 pandemic had on their personal lives, and domestic violence. Moreover, we also made efforts to determine the impact that domestic violence has on mental health and analyze the perceptions of respondents with regard to mental health help seeking.

Data Collection
Although there were more than 35,000 people in the participating groups and organizations, a total of only 136 individuals voluntarily participated in this study by filling in the web-based questionnaire. We assumed that at least 1% of the questionnaire recipients would be interested in participating in this study and were expecting to have more than 350 respondents. However, we had a much lower response rate than what we initially assumed. Sociocultural stigma and the burden of the increased number of web-based surveys being conducted during the study period by different organizations might have had an impact on the response rate.

Data Analysis Process
All information was gathered via Google Forms and was recoded into variables. The data were reviewed and sorted before starting the analysis. Data were analyzed by using IBM SPSS (version 22.0; IBM Corporation). The analysis techniques conducted were mainly for analyzing descriptive statistics.

Ethical Consideration
This study was approved by the National Institute of Mental Health, Dhaka, and conformed to the provisions of the Declaration of Helsinki. All ethical procedures were maintained and followed during this study, including the process of maintaining web-based data privacy and security for sensitive data.

Results

We analyzed 136 respondents aged between 17 and 50 years; their mean age was 24.26 years (SD 5.15 years). The sociodemographic characteristics of the respondents are reported in Table 1.

A total of 36.8% (50/136) of the respondents reported that they were survivors of domestic violence, and this prevalence rate was about 3 times higher among females (36/50, 72%) than that among males (14/50, 28%). Of the 50 domestic violence survivors, 16 (32%) experienced domestic violence very often, 10 (20%) sometimes experienced domestic violence, and 24 (48%) rarely experienced domestic violence. The most common type of violence that the respondents faced was mental abuse (n=34, 65.4%; Table 2). However, one should also consider that 3.7% of the respondents were unwilling to answer the question regarding whether they were survivors of domestic violence. Moreover, 24.2% (33/136) of respondents (male: 21.2%; female: 78.8%) experienced domestic violence for the first time during the lockdown period, and this violence was perpetrated by their family members. Further, 5.1% (n=7) of domestic violence survivors faced such violence very often during the lockdown period. In addition, 22.8% (31/136) of the participants revealed that their other family members also experienced domestic violence, and 37.5% (51/136) of respondents came to know that other relatives or friends experienced domestic violence during the lockdown period.

In total, 41.2% of participants reported that they faced a mild economic crisis, and 30.1% of participants reported that they faced a significant economic crisis. Table 2 shows that among the different forms of abuse, the prevalence of economical abuse almost tripled during the lockdown period. Furthermore, a positive correlation was observed between changes in economic conditions after the COVID-19 lockdown started and experiencing domestic violence for the first time after the COVID-19 pandemic started (n=134; r=0.107); however, the correlation was not statistically significant (P=.89).

The participants had different opinions regarding the reasons behind domestic violence during the COVID-19 pandemic. These opinions are shown in Table 3.

Our study also revealed that about 45% of the respondents experienced web-based sexual harassment in their lifetime. We also found that social media engagement was linked to respondents’ experiences of domestic violence. More than 96% (102/136, 96.3%) of the respondents believed that they needed mental health support. However, 75% (102/136) of the respondents did not know how to avail such services in Bangladesh.
Table 1. Sociodemographic characteristics of the respondents (N=136).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Respondents, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>35 (25.7)</td>
</tr>
<tr>
<td>Female</td>
<td>101 (74.3)</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
</tr>
<tr>
<td>Unmarried</td>
<td>103 (75.7)</td>
</tr>
<tr>
<td>Married</td>
<td>30 (22.1)</td>
</tr>
<tr>
<td>Separated</td>
<td>2 (1.5)</td>
</tr>
<tr>
<td>Widowed</td>
<td>1 (0.7)</td>
</tr>
<tr>
<td><strong>Occupation</strong></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>92 (67.6)</td>
</tr>
<tr>
<td>Service holder</td>
<td>26 (19.1)</td>
</tr>
<tr>
<td>Businessperson</td>
<td>5 (3.7)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>8 (5.9)</td>
</tr>
<tr>
<td>Other</td>
<td>5 (3.7)</td>
</tr>
<tr>
<td><strong>Family status</strong></td>
<td></td>
</tr>
<tr>
<td>Nuclear family</td>
<td>110 (80.9)</td>
</tr>
<tr>
<td>Joint family</td>
<td>20 (14.7)</td>
</tr>
<tr>
<td>Currently staying out of family</td>
<td>6 (4.4)</td>
</tr>
</tbody>
</table>

Table 2. Distribution of the different types of violence before the lockdown period and after the lockdown was imposed.

<table>
<thead>
<tr>
<th>Types of violence</th>
<th>Before lockdown, n (%)</th>
<th>After the first lockdown was imposed, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical</td>
<td>10 (19.2)</td>
<td>4 (10.5)</td>
</tr>
<tr>
<td>Psychological</td>
<td>34 (65.2)</td>
<td>26 (68.4)</td>
</tr>
<tr>
<td>Sexual</td>
<td>1 (1.9)</td>
<td>1 (2.6)</td>
</tr>
<tr>
<td>Economical</td>
<td>3 (5.8)</td>
<td>6 (15.8)</td>
</tr>
<tr>
<td>Other harmful traditional practices</td>
<td>4 (7.7)</td>
<td>1 (2.6)</td>
</tr>
</tbody>
</table>

Table 3. The possible reasons behind the increased prevalence of domestic violence.

<table>
<thead>
<tr>
<th>Reasons</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failure to manage increasing stress</td>
<td>35 (25.7)</td>
</tr>
<tr>
<td>Deterioration of economic status</td>
<td>26 (19.1)</td>
</tr>
<tr>
<td>Increased duration of stay</td>
<td>26 (19.1)</td>
</tr>
<tr>
<td>Fear of getting infected and panic</td>
<td>17 (12.5)</td>
</tr>
<tr>
<td>Moral decay and family learning</td>
<td>12 (8.8)</td>
</tr>
<tr>
<td>Patriarchy</td>
<td>10 (7.4)</td>
</tr>
<tr>
<td>Deteriorating mental health and preexisting mental illness</td>
<td>7 (5.1)</td>
</tr>
<tr>
<td>Other</td>
<td>3 (2.2)</td>
</tr>
</tbody>
</table>

Discussion

Violence against women is a major public health issue, and managing this issue is a key objective of the sustainable development goals of Bangladesh. However, during the COVID-19 crisis, the incidence of domestic violence has supposedly increased, and China, France, Italy, Brazil, and Spain have reported the increased incidence of domestic violence during the COVID-19 crisis [14]. There is a significantly higher number of females reporting that they have experienced domestic violence compared to this number among males. A male-dominated society or patriarchal society, like
that in Bangladesh, provides more scope for females to experience domestic violence [1]. Other factors that participants reported as the causes of violence were financial instability, mental stress, and the increased number of opportunities resulting from the COVID-19 lockdown. Moreover, it has also been observed that pandemics and any other crisis situations increase the risk of abuse among vulnerable populations [8]. However, the number of males reporting to be survivors of domestic violence is also noteworthy for a male-dominated country like Bangladesh, and this topic demands further research.

The COVID-19 pandemic has increased the risk of domestic violence, as survivors have had to remain with their abusive family members for a longer period of time [15,16]. It has been observed that mothers-in-law and sisters-in-law act as instigators in domestic violence incidents in Bangladesh, and this finding is consistent with those of other qualitative studies in India [17,18]. Therefore, the COVID-19 lockdown has provided more opportunities for intimate partners and other abusive family members to engage in violent activities. As those who experience domestic violence are at risk of developing various mental and physical conditions [19], a responsive reporting system, along with effective mental health care, is of utmost importance to domestic violence survivors [20]. Moreover, it is necessary to inform the survivors of domestic violence about the available mental health support services in a country. Even if these survivors know about the different hospitals in a country, they fail to avail such services due to geographic distance, the lack of freedom to seek such services, the fear of experiencing further abuse from their families, and the need to provide the cost of such services. These barriers can be minimized by telepsychiatry or e-mental health services. Telepsychiatry (ie, the long-distance provision of psychiatry services) emerged with the promising effect of overcoming potential barriers and improving access to mental health services for all [19,21].

The role of telepsychiatry has become more important during the COVID-19 crisis in Bangladesh, as telepsychiatry services provide psychological support to the those who experience gender-based violence. Telepsychiatry support can provide more anonymity by providing the option of communicating with professionals 24-7 through audio, video, or chat platforms, thereby ensuring that that the limited number of professionals serve the highest possible number of people. Domestic violence survivors need to have the phone numbers of national call centers to obtain different kinds of support. They also need different web-based and mobile-based mental health services, such as MonerDaktar [22], MindTale, and Maya Apa; these services have made remarkable contributions by providing mental health support. Governments and policy makers should consider how they can use technology-based psychiatry services or digital psychiatry services to serve the thousands of women who are facing different forms of violence every day and need mental health support.

**Conclusion**

The prevalence of domestic violence has been increasing in Bangladesh during the COVID-19 crisis, and this crisis will persist for a few more months or years. As such, mental health burden resulting from domestic violence will rise in prevalence. We need a multilayered holistic plan for supporting and providing cost-effective, high-quality mental health support for domestic violence survivors.

**Acknowledgments**

We are thankful to Shahana Parveen, Ashiqur Rahman, Aminul Islam, Sharmin Ara, Dr Nowreen Ahmed, and Nadia Alam for helping us with this endeavor.

**Conflicts of Interest**

None declared.

**References**


https://formative.jmir.org/2021/9/e24624


©Tanjir Rashid Soron, Md Ashiqur Rahman Ashiq, Marzia Al-Hakeem, Zaid Farzan Chowdhury, Helal Uddin Ahmed, Chaman Afrooz Chowdhury. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.09.2021. 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.
Early Detection of Symptom Exacerbation in Patients With SARS-CoV-2 Infection Using the Fitbit Charge 3 (DEXTERITY): Pilot Evaluation

Kan Yamagami1*, MD; Akihiro Nomura1*, PhD; Mitsuhiro Kometani2, PhD; Masaya Shimojima1, PhD; Kenji Sakata1, PhD; Soichiro Usui1, PhD; Kenji Furukawa3, PhD; Masayuki Takamura1, PhD, Prof Dr; Masaki Okajima4, PhD, Prof Dr; Kazuyoshi Watanabe5, MD; Takashi Yoneda2, PhD, Prof Dr

1Department of Cardiovascular Medicine, Graduate School of Medical Sciences, Kanazawa University, Kanazawa, Japan
2Department of Health Promotion and Medicine of the Future, Kanazawa University, Kanazawa, Japan
3Health Care Center, Japan Advanced Institute of Science and Technology, Ishikawa, Japan
4Intensive Care Unit, Kanazawa University Hospital, Kanazawa, Japan
5Japan Community Health Care Organization Kanazawa Hospital, Kanazawa, Japan
*these authors contributed equally

Corresponding Author:
Akihiro Nomura, PhD
Department of Cardiovascular Medicine
Graduate School of Medical Sciences
Kanazawa University
13-1 Takaramachi
Kanazawa
Japan
Phone: 81 076 265 2000
Email: anomura@med.kanazawa-u.ac.jp

Abstract

Background: Some patients with COVID-19 experienced sudden death due to rapid symptom deterioration. Thus, it is important to predict COVID-19 symptom exacerbation at an early stage prior to increasing severity in patients. Patients with COVID-19 could experience a unique “silent hypoxia” at an early stage of the infection when they are apparently asymptomatic, but with rather low SpO2 (oxygen saturation) levels. In order to continuously monitor SpO2 in daily life, a high-performance wearable device, such as the Apple Watch or Fitbit, has become commercially available to monitor several biometric data including steps, resting heart rate (RHR), physical activity, sleep quality, and estimated oxygen variation (EOV).

Objective: This study aimed to test whether EOV measured by the wearable device Fitbit can predict COVID-19 symptom exacerbation.

Methods: We recruited patients with COVID-19 from August to November 2020. Patients were asked to wear the Fitbit for 30 days, and biometric data including EOV and RHR were extracted. EOV is a relative physiological measure that reflects users’ SpO2 levels during sleep. We defined a high EOV signal as a patient’s oxygen level exhibiting a significant dip and recovery within the index period, and a high RHR signal as daily RHR exceeding 5 beats per day compared with the minimum RHR of each patient in the study period. We defined successful prediction as the appearance of those signals within 2 days before the onset of the primary outcome. The primary outcome was the composite of deaths of all causes, use of extracorporeal membrane oxygenation, use of mechanical ventilation, oxygenation, and exacerbation of COVID-19 symptoms, irrespective of readmission. We also assessed each outcome individually as secondary outcomes. We made weekly phone calls to discharged patients to check on their symptoms.

Results: We enrolled 23 patients with COVID-19 diagnosed by a positive SARS-CoV-2 polymerase chain reaction test. The patients had a mean age of 50.9 (SD 20) years, and 70% (n=16) were female. Each patient wore the Fitbit for 30 days. COVID-19 symptom exacerbation occurred in 6 (26%) patients. We were successful in predicting exacerbation using EOV signals in 4 out of 5 cases (sensitivity=80%, specificity=90%), whereas the sensitivity and specificity of high RHR signals were 50% and 80%,
respecitively, both lower than those of high EOV signals. Coincidental obstructive sleep apnea syndrome confirmed by polysomnography was detected in 1 patient via consistently high EOV signals.

**Conclusions:** This pilot study successfully detected early COVID-19 symptom exacerbation by measuring EOV, which may help to identify the early signs of COVID-19 exacerbation.

**Trial Registration:** University Hospital Medical Information Network Clinical Trials Registry UMIN000041421; https://upload.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000047290

**KEYWORDS**
COVID-19; silent hypoxia; wearable device; Fitbit; estimated oxygen variation; detection; infectious disease; pilot study; symptom; outpatient; oxygen; sleep; wearable

**Introduction**

The COVID-19 pandemic caused by SARS-CoV-2 has resulted in over 168 million cases and 3.5 million deaths worldwide as of late May 2021 [1]. The virus, with a long latency period of 2 to 14 days from initial infection to the onset of symptoms, is most transmissible just before symptoms appear [2]. Approximately half of infected patients are asymptomatic in some areas [3] but can spread the infection to others. Therefore, being asymptomatic might be one of the leading causes behind the global spread of the disease [4].

Some patients with COVID-19 experience rapid deterioration after 1 week of initial symptom onset, requiring oxygen or care in the intensive care unit (ventilators and extracorporeal membrane oxygenation [ECMO]) for severe pneumonia and acute respiratory distress syndrome–like symptoms [5]. According to a New York City report, the mortality rate of COVID-19 for patients on ventilators is over 75% [6]. The benefit of antiviral medication such as remdesivir or dexamethasone might be most apparent when it is used before symptom exacerbation [7,8]. Thus, it is important to predict COVID-19 symptom exacerbation at an early stage before patients experience increasing severity. In Japan, an increasing number of patients with COVID-19 with mild symptoms are managed in their homes or hotels to make effective use of medical resources [9]. However, in some cases, worsening symptoms resulted in death before the patient could be admitted to a hospital [10,11]. Therefore, it is necessary to construct an alarm system to detect signs of severity beforehand to prevent patients from serious illness or death while waiting at home or in hotels.

To evaluate the severity of pulmonary diseases, blood oxygen saturation ($\text{SpO}_2$) levels, measured by pulse oximetry, usually provides important information [12]. In fact, patients with COVID-19 could experience a unique “silent hypoxia” at an early stage of the infection when they are apparently asymptomatic, but have rather low $\text{SpO}_2$ levels [13]. Since a low $\text{SpO}_2$ normally indicates a severe pulmonary reaction to the disease, monitoring $\text{SpO}_2$ could provide potential biometric data to predict impending disease deterioration [14]. Indeed, low $\text{SpO}_2$ levels could predict future disease exacerbation in patients with chronic obstructive pulmonary disease [12]. However, except in special situations in an intensive care unit, continuous $\text{SpO}_2$ monitoring in daily life is unlikely because the measuring equipment normally needs to be clipped to one’s finger for every measurement. Recently, high-performance wearable devices, such as the Apple Watch and Fitbit, have become commercially available to monitor biometric data including steps, resting heart rate (RHR), physical activity, sleep quality, and even estimated oxygen variations (EOV; a relative physiological measure of a user’s $\text{SpO}_2$ levels during sleep). For example, one study used biometric data obtained from a Fitbit, a smartwatch device worn on one’s wrist, and showed that increased RHR and decreased sleep duration were associated with flu-like symptoms [15]. With COVID-19, some studies have demonstrated that changes in certain biometric indicators, including RHR, sleep duration, or respiratory rate, from the baseline might predict the occurrence of COVID-19 symptoms before their onset among those who use wearable devices on a daily basis [16-18]. However, it is still unclear whether the detection of variations in blood oxygen level might be useful for predicting COVID-19 severity by wearable devices, or whether wearable devices can detect signs of symptom exacerbation in patients with a confirmed case of COVID-19.

Here, we conducted the DEXTERITY pilot study, leveraging a wearable device to obtain biometric data, EOV and RHR, in particular, in patients diagnosed with COVID-19 to predict symptom exacerbation.

**Methods**

**Participants**

We prospectively recruited 28 patients with a positive SARS-CoV-2 polymerase chain reaction (PCR) test, according to the study protocol approved by the Kanazawa University and the Japan Community Health Care Organization (JCHO) Kanazawa Hospital Institutional Review Boards. This study was conducted from August to November 2020 at JCHO Kanazawa Hospital in Kanazawa, Japan. We performed the study in compliance with the Ethical Guidelines for Medical and Health Research Involving Human Subjects, the Declaration of Helsinki, and other guidelines in Japan. We registered this study with the University Medical Information Network Clinical Trial Registry on August 14, 2020 (UMIN000041421).

We included patients who were diagnosed with COVID-19 and had a positive SARS-CoV-2 PCR test result within 1 week before enrollment in the study. We excluded patients who met
the following criteria: (1) unable to wear and use the wearable device, (2) unable to connect the wearable device to the smartphone app, (3) unable to download or use the smartphone app, and (4) unable to provide informed consent because of severe COVID-19 symptoms. We obtained informed consent electronically via mobile platforms using the Research Electronic Data Capture (REDCap) system from all participants.

Wearable Device and Data Extraction

We provided a Fitbit Charge 3 to each participant. They wore the wearable device for 30 days to detect COVID-19 symptom exacerbation, which normally occurs at 7 to 14 days after the onset of initial symptoms [5]. The Fitbit Charge 3 was connected to each patient’s smartphone via the Fitbit app, and their biometric data, including RHR and sleep quality, were extracted through the Fitabase, a web-based Fitbit-derived data extraction system for clinical studies.

Patients were asked to complete electronic questionnaires in the REDCap system, which included questions on baseline characteristics (age; sex; height; weight; BMI; smoking status [current, former, or never]; presence of hypertension, diabetes mellitus, or dyslipidemia; any other medical history, and any medication use) at the time of enrollment. Additionally, we obtained information regarding COVID-19–related symptoms from the patients during hospitalization and after discharge over the course of the 30-day study period. COVID-19–related symptoms included fever, cough, fatigue, difficulty in breathing, nausea, diarrhea, dysosmia, or dysgeusia. We also conducted weekly checks via telephone on symptom improvement or exacerbation, readmission to the hospital, oxygenation, use of mechanical ventilation, use of ECMO, or death resulting from all causes.

Estimated Oxygen Variation

In addition to the biometric data obtained from the Fitbit Charge 3, we directly collected daily EOV graphs by taking screenshots of each patient’s app interface. EOV is internally calculated using an algorithm that estimates the variation in the reflected rate from the reflected optical signals every minute. If a patient’s oxygen level is stable, the variation is low or close to zero. However, if a patient’s oxygen level exhibits a significant dip and recovery within the index period, the variation shows a high signal. Variations that cross the threshold line are shown in Figure 1. We defined a high EOV (single day) signal as an EOV value that passes the threshold line on the graph at one or more times during sleep. Since we considered that a symptom deterioration signal could last several days, if a high single-day EOV signal continues for 2 or more consecutive days, we regarded the signal as a high EOV signal and calculated the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

Figure 1. Representative graphs of low (negative) and high (positive) estimated oxygen variation (EOV) signals from the Fitbit Charge 3.
Outcomes
The primary outcome was the composite of deaths by all causes, use of ECMO, use of mechanical ventilation, oxygenation, and exacerbation of COVID-19 symptoms, irrespective of readmission. We also assessed each outcome individually as secondary outcomes. We made phone calls to each patient every week and asked whether their symptoms were stable or had changed. If the patient’s symptoms changed, we asked when it occurred, and 2 study investigators judged if the change was deemed to be an exacerbation or not. We defined exacerbation of COVID-19 as fever, dyspnea or intense malaise, and other common cold symptoms such as cough, gastrointestinal symptoms, or symptoms considered similar to those seen when the patient had COVID-19. If the patient did not answer the call, we contacted the patient’s family instead and instructed them to ask the patient to pick up the phone. If that did not work, we called on a different day of the week.

Statistical Analysis
The baseline profile is shown as the mean (SD) or median with quantiles (for continuous variables), or proportions (for categorical variables). We overlayed and compared the onset of the outcomes and the number of days high EOV signals were detected. We defined a successful prediction of an outcome by EOV as the presence of high EOV signal(s) within 2 days of the onset of the outcome. We also defined high RHR signals as a daily RHR exceeding 5 beats/day compared with the minimum RHR of each patient during the study period. We calculated the sensitivity, specificity, PPV, and NPV for both the high EOV signal and the high RHR signal for primary outcome prediction. Sensitivity was defined as the number of true positives divided by the number of exacerbation events, where a true-positive event refers to the appearance of a high EOV signal and exacerbated symptoms. Specificity was defined as the number of true negatives divided by the number of events without exacerbated symptoms, where a true negative refers to an event where a high EOV signal does not appear and the patient’s symptom does not exacerbate. PPV was defined as the number of true positives divided by the number of high EOV signals. NPV was defined as the number of true negatives divided by the number of EOV signals that are not high. All tests were two-sided, and significant differences were considered when \( P < 0.05 \). We used R, version 3.6.1 (R Foundation for Statistical Computing) for the analyses.

Data and Code Availability
The anonymized data set of this study will be available from the corresponding author upon publication. The investigator may only use the data for the purpose outlined in the request. Data redistribution is prohibited.

Custom codes or mathematical algorithms were not used in this study.

Results
Study Participants and Baseline Characteristics
During the study period, 43 patients were admitted to JCHO Kanazawa Hospital. Of these, 15 were excluded from the study; 8 did not consent to join the study; 6 did not have smartphones or could not download the Fitbit app; and 1 had a severe respiratory condition and was immediately transferred to another hospital. Thus, we prospectively recruited 28 SARS-CoV-2 PCR-positive patients in this pilot study. Of this sample, 4 patients could not connect to the Fitbit account, and 1 patient had no personal email address. Therefore, 23 patients were followed up for 30 days and included in further analyses (Figure 2).

Table 1 shows the baseline characteristics of the patients. A total of 23 patients were included. The patients had a mean age of 50.9 (SD 20) years, and 70% (n=16) were female. Two (9%) patients had a history of malignancy (2 patients with breast cancer), but none had cardiovascular or cerebrovascular diseases. Symptoms at admission were dyspnea in 2 (9%) patients, fever
in 12 (52%) patients, dysgeusia in 3 (13%) patients, dysosmia in 1 (4%) patient, sore throat in 1 (4%) patient, and no symptoms in 4 (17%) patients. The median interval from initial COVID-19 symptoms to Fitbit use was 5 days (range 1-9 days), and the median days of wearing the Fitbit was 19 (IQR 15.5-28) days.

Table 1. Baseline characteristics.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Participants (N=23)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>50.9 (20)</td>
</tr>
<tr>
<td>Gender (female), n (%)</td>
<td>16 (70)</td>
</tr>
<tr>
<td>Body weight (kg), mean (SD)</td>
<td>58.7 (16)</td>
</tr>
<tr>
<td>BMI (kg/m²), mean (SD)</td>
<td>22.8 (4.7)</td>
</tr>
<tr>
<td>Comorbidities</td>
<td></td>
</tr>
<tr>
<td>Hypertension, n (%)</td>
<td>5 (22)</td>
</tr>
<tr>
<td>Diabetes mellitus, n (%)</td>
<td>2 (9)</td>
</tr>
<tr>
<td>Dyslipidemia, n (%)</td>
<td>5 (22)</td>
</tr>
<tr>
<td>Medical history, n (%)</td>
<td></td>
</tr>
<tr>
<td>Malignancy</td>
<td>2 (9)</td>
</tr>
<tr>
<td>Cardio- or cerebrovascular diseases</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Smoking status, n (%)</td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>12 (52)</td>
</tr>
<tr>
<td>Former</td>
<td>9 (39)</td>
</tr>
<tr>
<td>Current</td>
<td>2 (9)</td>
</tr>
<tr>
<td>Symptoms at admission, n (%)</td>
<td></td>
</tr>
<tr>
<td>Fever</td>
<td>12 (52)</td>
</tr>
<tr>
<td>Dysgeusia</td>
<td>3 (13)</td>
</tr>
<tr>
<td>Dyspnea</td>
<td>2 (9)</td>
</tr>
<tr>
<td>Dysosmia</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Sore throat</td>
<td>1 (4)</td>
</tr>
</tbody>
</table>

Estimated Oxygen Variation and Outcomes

Figure 3 demonstrates a summary of high EOV signals among the 23 patients. We observed 48 high EOV signals (73 single-day high signals) during the study. We found a median of 1 high EOV signal per patient (IQR 1-3). The median percentage of EOV per day was 16% (IQR 11%-19%). Of the 23 patients, we excluded 1 patient (JCHO-023) from further analyses because of obstructive sleep apnea syndrome (OSAS) detected via a polysomnography (the details are described in the “Representative Case 3: Obstructive Sleep Apnea Syndrome (JCHO-023)” section).

The primary outcomes (symptom exacerbation events) occurred 7 times in 6 patients during the study period. The primary outcomes included the use of high-flow nasal cannula (HFNC) (n=2), exacerbation of cough (n=2; 1 patient was readmitted to the hospital), exacerbation of dysosmia (n=1), and experience of fever and general malaise (n=1). In patients with EOV data, we successfully observed high EOV signals within 2 days of the symptom exacerbation events in 4 out of 5 cases (sensitivity=80%, specificity=90%), although NPV was 99.7% and PPV was only 9.3%. The reference sensitivity, specificity, PPV, and NPV of high RHR signals for detecting these events were 50%, 88%, 6.7%, and 99.1%, respectively, all of which were lower than those of high EOV signals. The clinical course of the patients with COVID-19 are shown in Multimedia Appendix 1.

Next, we reported representative cases in whom we could successfully observe (JCHO-008) or not observe (JCHO-016) high EOV signals just before the exacerbation of COVID-19 symptoms (Figure 4). In addition, we presented a patient with COVID-19 (JCHO-023) in whom we unintentionally detected OSAS due to the extreme and consistently high EOV signals observed during the study period.
Figure 3. Summary of high estimated oxygen variation (EOV) signals and events among patients with a positive SARS-CoV-2 polymerase chain reaction test. The vertical columns represent the ID of each patient, and the horizontal axis represents the number of days of wearing a Fitbit. Blue squares represent a single-day high EOV signal. Gray columns denote the days when no EOV data were obtained. Red columns indicate the days that primary outcomes (symptom exacerbation events) occurred. JCHO: Japan Community Health Care Organization.

Figure 4. The clinical courses of patients JCHO-008, JCHO-016, and JCHO-023. The vertical columns represent estimated oxygen variation (EOV), resting heart rate (RHR), body temperature (BT), and cough or fatigue severity levels. The horizontal axis represents the number of days after a positive SARS-CoV-2 polymerase chain reaction (PCR) test (JCHO-008) or the occurrence of COVID-19–related symptoms (JCHO-016 and JCHO-023). Gray columns represent a day when no EOV data were obtained. FiO₂: fraction of inspired oxygen; JCHO: Japan Community Health Care Organization; NHF: nasal high flow; Re-ad: readmission.
Representative Case 1: Successfully Detected (JCHO-008)
A 65-year-old woman with a history of gastric cancer was referred to our hospital due to a positive SARS-CoV-2 PCR test result (Figure 4A). She was asymptomatic at the time of diagnosis by PCR. However, she began coughing just before hospital admission, and her SpO₂ was 98% at room air. Although she experienced a moderate fever (up to 38.8°C) with computed tomography (CT)—confirmed pneumonia for 1 week, she became afebrile and asymptomatic except for a slight cough on the day of discharge on day 12. A high EOV signal was once detected shortly after discharge, but no symptom exacerbation occurred after the signal. However, another high EOV signal occurred on day 25. Since her cough suddenly worsened following the fever after the signal, she visited a hospital on day 30. Her chest CT exhibited COVID-19–like infiltration and interstitial shadow, and she was readmitted to the hospital and diagnosed with COVID-19–induced pneumonia recurrence. In this case, the high EOV signal on day 25 was successfully observed just before the symptom exacerbation on day 26.

Representative Case 2: Undetected (JCHO-016)
There was an undetected case of EOV for COVID-19 symptom exacerbation before symptom onset. A 75-year-old man with a history of hypertension was referred to our hospital due to a positive SARS-CoV-2 PCR test (Figure 4B). At admission, his body temperature was 39.6°C, he had an SpO₂ of 98% (via O₂ nasal canula at a rate of 2 L/min), and his chest CT showed no shadow compatible with COVID-19. However, his SpO₂ gradually worsened, and he had to use HFNC to maintain oxygenation on day 9 (the maximum fraction of inspired oxygen [FiO₂] was 70%). Chest CT on day 8 revealed ground-glass opacity compatible with COVID-19. During this exacerbation, no high EOV signals were observed. However, just after HFNC discontinuation on day 13, high EOV signals were observed daily from days 14 to 17, but no exacerbation of symptoms was found afterward. Chest CT on day 15 showed remarkable improvement of ground-glass opacity. He discontinued oxygenation on day 21 and was discharged from the hospital on day 28. After discharge, high EOV signals were detected twice, but again, no exacerbations of any symptoms were found after these signals.

Representative Case 3: Obstructive Sleep Apnea Syndrome (JCHO-023)
There was a patient with COVID-19 in whom we coincidentally detected OSAS due to the consistently high EOV signals observed during the study period. A 53-year-old man with a history of hypertension was referred to our hospital due to a positive SARS-CoV-2 PCR test (Figure 4C). He experienced a mild fever (up to 37.8°C) and a moderate cough (4 days before admission). At admission, his body temperature was 38.4°C, his SpO₂ level was 96% (O₂ nasal canula, 2 L/min), and his chest CT showed shadows compatible with COVID-19. Just after admission, high EOV signals were observed from days 1 to 4, but no exacerbation of symptoms was found after these signals. His body temperature and cough improved gradually after admission, and he became afebrile and asymptomatic after day 9. His SpO₂ also improved gradually, and he discontinued oxygenation on day 11. High EOV signals were observed on day 10, days 12 to 13, and day 16, but no symptom exacerbation occurred after these signals. He was discharged from the hospital on day 17 and was afebrile and asymptomatic thereafter. However, a long-lasting high EOV signal was observed from days 18 to 29 without any symptoms. At this time, we assumed that he might be having sleep apnea syndrome (SAS). He underwent polysomnography by a portable polysomnogram monitor (SAS-2200, Nihon Kohden). His 3% oxygen desaturation index was 33.6 per hour, indicating severe OSAS. In this case, the high EOV signal was observed not due to the exacerbation of COVID-19 symptoms but by the existing condition of OSAS.

Discussion
Principal Findings
This is the first prospective pilot study to assess whether EOV, a relative physiological measure that indicates continuous SpO₂ variations during sleep, using a Fitbit wearable device, could predict early exacerbation signs of SARS-CoV-2 infection before their onset. We demonstrated that the high EOV signals observed just before symptom exacerbation in 4 out of 5 cases (80%) was higher than that in RHR signals. In addition, we detected a severe case of OSAS via the intermittently high EOV signals obtained from the Fitbit device.

This study yielded several important findings. First, the high EOV signal provided by the Fitbit demonstrated a favorable sensitivity (80%) and high NPV (99.7%) (both higher than those of RHR signals) for COVID-19 symptom exacerbations prior to their onset. The high sensitivity and NPV of the device and signal used in this study are of particular importance for screening and early diagnosis of COVID-19 exacerbations, which could accurately identify cases that warrant closer inpatient or outpatient monitoring. In some patients with COVID-19, silent hypoxia has been reported, with remarkably low SpO₂ levels, while having minimal typical symptoms such as fever, cough, or fatigue [19]. Following silent hypoxia, patients experienced apparent symptoms [20,21]. The mechanism of silent hypoxia involves a combination of many factors, including the response of the respiratory centers and the effect of comorbidities (eg, diabetes mellitus) and older age on breathing control [13]. Additionally, the idiosyncratic action of the coronavirus on receptors involved in chemosensitivity to oxygen has been demonstrated before [13]. Indeed, angiotensin-converting enzyme 2, the cell receptor of SARS-CoV-2, is expressed in the carotid body, the site at which the chemoreceptors sense oxygen [22]. The development of a thrombi within the pulmonary vasculature may also be related to silent hypoxia [21]. In this study, since silent hypoxia–like abrupt SpO₂ depletion without apparent symptoms occurred in some patients, the EOV could successfully predict the exacerbation of COVID-19 symptoms. Although RHR is inferior to EOV in terms of sensitivity, it is not data that should be discarded, and there is a possibility that the prediction accuracy can be further improved by creating an index comprising both EOV and RHR.
Second, although high EOV signals showed high sensitivity and NPV for detecting COVID-19 symptom exacerbation, the PPV of high EOV signals was only 9.3%. High EOV signals may be invoked not only by SpO₂ exacerbation but also by other situations including alcohol intake, emotional stress events, or medications, as heart rate increases in such situations [16]. Recent studies that aimed to predict the onset of COVID-19 showed that RHR- and sleep duration-derived indices tended to yield more false-positives among patients who had been diagnosed with COVID-19 [16,18]. The same situation might also have occurred for EOV. We still need to take into account the multiple factors that could interfere with a high EOV signal’s ability to predict COVID-19 symptom exacerbation.

Third, we detected a case of severe OSAS in a patient by chance due to consistently high EOV signals. To our knowledge, this is the first study to report a clinical case of extremely high EOV signals obtained by the Fitbit to detect SAS. SAS is a common disorder that causes patients to temporarily stop or decrease their breathing repeatedly during sleep [23]. It is caused by a dynamic upper airway collapse and results in low SpO₂ during the night [24]. We believe that these SpO₂ depletion events were detected by the Fitbit as high EOV signals. At present, Fitbit Inc is applying to the United States Food and Drug Administration to include this EOV function in a medical device to diagnose SAS. Once approved, Fitbit may soon become an innovative device to diagnose SAS.

Limitations
This study has several limitations. First, the Fitbit wearable device, including its functions (eg, EOV) and analysis algorithms, was not yet approved as a medical device at the time of this study. Second, we could not acquire sufficient baseline data for each patient regarding RHR, EOV, and other biometric data because the duration of the study period over which the patients wore the Fitbit was only 30 days. Baseline biometric data for each physiological metric are very important to distinguish abnormal signals from normal variations at the individual level. In this study, we set the minimum value of each biometric factor during the study as the baseline value, but this could have affected the evaluation of positive signals for each metric. Third, we could not assess sleep duration as one of the biometric markers for exacerbation detection. We attempted to investigate the association of each sleep duration per day with the exacerbation of COVID-19 symptoms, but had to abandon it due to a lack of baseline data and significant variations between the in-hospital and after-discharge periods. Fourth, elderly patients who could not use a smartphone daily did not participate in the study, although they were more likely to have exacerbated SARS-CoV-2 infection symptoms. This problem can be solved by having medical staff operate the older patients’ smartphones instead under adequate infection protection. This, however, is difficult to do in a busy hospital ward. Fifth, the number of people in whom the primary outcome occurred (ie, 5) is too small to be generalizable. However, this is a pilot study, and future studies with a larger sample size will be needed to validate our results.

Conclusions
In conclusion, we demonstrated that EOV from the Fitbit wearable device could detect 80% of symptom exacerbations among patients with SARS-CoV-2 infection before their onset. Additionally, we coincidentally detected OSAS through consistently high EOV signals. In the future, we hope to integrate EOV and other physiological metrics such as RHR, respiratory rate, or sleep data to improve the prediction accuracy of COVID-19 symptom exacerbations in advance.

Acknowledgments
We thank all the participants, the medical staff, and the COVID-19 Special Assistant Team (COVSAT) at Kanazawa University Hospital, as well as Shinichiro Takashima, Hirofumi Okada, Satoshi Niwa, Hayato Tada, Toyonobu Tsuda, and the many others at Kanazawa University Hospital and Japan Community Health Care Organization (JCHO) Kanazawa Hospital for their outstanding work against COVID-19 while simultaneously cooperating with this study. Additionally, we express our gratitude to Conor Heneghan and Tomoyasu Chigahara for their kind support regarding the EOV analysis. Moreover, we appreciate the help received from Jia Yu and Yasuhiro Imai for REDCap maintenance, electronic consent production, data cleaning, and data management. The investigational device used in this study, the Fitbit Charge 3, was provided by Fitbit Japan.

Authors' Contributions
KY and AN wrote the draft of the manuscript and have access to all the study data. AN takes responsibility for data integrity and analyzed the data. KY, AN, MK, MS, KS, SU, KF, MT, MO, and TY conceptualized and designed the study. KY, AN, and KW collected clinical data from the participants. All the authors approved the final version of the manuscript.

Conflicts of Interest
AN received consulting fees from CureApp, Inc. AN was a cofounder of the CureApp Institute. KF declares lecture fees from Sanofi KK and Eli Lilly Japan KK. The other authors declare no competing interests.

Multimedia Appendix 1
The clinical course of the patients with COVID-19.
[ PNG File , 653 KB - formative_v5i9e30819_app1.png ]
References


Abbreviations

CT: computed tomography  
ECMO: extracorporeal membrane oxygenation  
EOV: estimated oxygen variation  
FiO₂: fraction of inspired oxygen  
HFNC: high-flow nasal cannula  
JCHO: Japan Community Health Care Organization  
NPV: negative predictive value  
OSAS: obstructive sleep apnea syndrome  
PCR: polymerase chain reaction  
PPV: positive predictive value  
REDCap: Research Electronic Data Capture  
RHR: resting heart rate  
SAS: sleep apnea syndrome  
SpO₂: oxygen saturation

©Kan Yamagami, Akihiro Nomura, Mitsuhiro Kometani, Masaya Shimojima, Kenji Sakata, Soichiro Usui, Kenji Furukawa, Masayuki Takamura, Masaki Okajima, Kazuyoshi Watanabe, Takashi Yoneda. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.09.2021. 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.